Investigating commitment flexibility in multi-agent contracts

Vanessa Teague
Department of Computer Science
Stanford University
vteague@cs.stanford.edu and Liz Sonenberg
Department of Information Systems
The University of Melbourne
l.sonenberg@dis.unimelb.edu.au

No Institute Given

Abstract. Reputation and commitment are important issues for automated contracting. Levelled commitment contracts, i.e. contracts where each party can decommit by paying a predetermined penalty, were introduced to allow self interested agents to accommodate events that unfolded since the contract was entered into. Various approaches to modelling reputation have been explored, allowing an agent to make decisions after considering not only current opportunities, but also possible longer term (social) implications of maintaining a reputation for reliability in contract fulfilment. An agent may consider both its own reputation for honesty and the past behaviour of others. We explore a combination of these issues in the context of a simple contracting scenario. Analytical and simulation results are presented.

Theme area: empirical evaluation

1 Introduction

The advent of widespread Electronic Commerce promises to have enormous benefits for consumers. Due to the proliferation of on-line auction houses such as eBay and on-line retailers such as Amazon.com, we now have a vast array of options for selling or purchasing almost anything. The resulting environment is well suited to intelligent autonomous agents, since some reasonably intelligent behaviour is required for bidding or bargaining but there are often too many possibilities for a human to investigate manually.

However, doing business on the internet also presents significant new problems. Possibly the most important of these is the problem of deciding when another entity on the web is trustworthy. Trustworthiness is a particularly important issue for organisers of on-line markets and auction houses, since the credibility of their market is undermined if participants frequently succeed in cheating others. The main question is whether market organisers should try to intervene and guarantee the trustworthiness of participants (as stock markets do, for example, by enforcing laws such as regular auditing of the traded companies), or adopt a *laissez faire* approach and assume that dishonourable parties will eventually acquire a bad reputation (as do eBay and MIT Media Lab's Kasbah system [7]).

The Internet is an environment that involves a large degree of uncertainty and changes quickly over time. In such environments, allowing agents to renegue on contracts sometimes can greatly improve performance, since agents can agree to contracts that seem beneficial and then decommit if circumstances change adversely.

However, trust of other agents is difficult to establish when some decommitment from contracts is permitted. If an agent enters into a contract, it needs to be confident that it will not be damaged by the dishonesty of other contract participants. In real-world (human) contracts, compliance can be enforced by an authoritative body such as a law court, but in artificial agent scenarios it is difficult or impossible to trace the behavior of an artificial agent to its human "owner". Therefore agents need some way of knowing how to evaluate the "trustworthiness" of others.

The model of trading we will adopt consists of a two-stage process, where the first stage involves only partial commitment. The second stage provides agents with an opportunity to leave before they commit fully. This has some intuitive appeal for an agent-mediated electronic commerce setting, since a normal pattern of behaviour would probably be to send an agent out on to the web to make tentative arrangements, but only spend large amounts of money with explicit confirmation from the human who sent it [8]. It also provides a precise way to measure "trustworthiness", since for any agent it is possible to count the number (or percentage) of times it honours its initial commitment.

Sandholm and colleagues introduce "Leveled Commitment Contracts" which allow agents to choose their level of commitment at the beginning [9–11]. Agents are allowed to decommit, but only if they pay the other participant(s) a predetermined penalty. The payment is assumed to be enforced by the market.

An example of a *laissez faire* market is is the work of Sen *et al.* [12, 13]. He assumes no market intervention, but that agents choose which others to work with according to a system of reciprocity. They remember which agents have reciprocated the favours they have done for them, gradually learning which others can be trusted. A stable community of cooperating (though self-interested) agents emerges.

The "brownie point" model of Grosz, Kraus and colleagues [4, 15] represents a moderate level of intervention, where there are both market-imposed penalties for decommitting and agent-imposed reluctance to do so.

1.1 Our problem

Our main interest was investigating different approaches to market intervention, the primary aims being to explore ones that elicit cooperative behaviour from the agents and to consider identifying characteristics of the environment and agent interactions that may make one approach more suitable than another.

We have devised a scenario general enough to incorporate all three of the decision factors described (decommitment penalties, brownie points and reliability) and flexible enough to adopt various combinations of their features. See Section 2.

Some parameterisations that involved only decommitment penalties could be analysed mathematically. In most cases, the analysis showed that the agents behaved optimally, in the sense that they always ended up maximising the total value of the targets they caught. In one case, agents were shown to behave "fairly", meaning that the payoff of the second-lowest-paid agent was maximised. Section 3 describes the analytical results.

Since it was not possible to analyse the "brownie point" and reciprocation based markets, we constructed a simulation in Java and performed experiments on the efficiency of the agents under different market conditions. The first experiments were mainly checks that the simulation reproduced the results contained in [1, 2] and [4]. These were generally successful, finding that a moderate emphasis on "brownie points" and small to medium decommitment penalties were optimal. Section 4 contains empirical investigations. The most interesting result was the formation of stable teams using the reliability-based approach (section 4.2). Other results confirmed the intuition that "brownie points" and decommitment penalties perform similar roles — section 4.4 shows that combining them does not increase performance compared to "brownie points" alone, and section 4.3 shows that "brownie points" perform similarly to decommitment penalties but slightly better. However, the reliability strategy was significantly different.

2 The Scenario

The theme is that agents band together to capture "targets". This corresponds to operating successfully together as a team. Agents are self-interested, though they may cooperate when they believe it is in their long-term interests.

There are a number of runs, each of which has four stages:

1. Each agent in turn chooses an initial target to try to capture. It moves to that target, and may make a commitment to stay there until the end of the run.
2. New targets appear.
3. Each agent in turn has one opportunity to leave its current target, pay any decommitment penalty required, and move to another.
4. (Clearing) The number of agents at each target is counted. If a target has "enough" agents it is considered captured and its value is counted towards the final total.

At the end of each round all targets disappear, all agents are released from all contracts, some new targets appear and all agents are removed from the board.

Each target has three important properties: the number of agents required to capture it, the amount of "money" it will pay each of them, and the decommitment penalties it will demand if they renegue. Suppose target k pays the first agent to arrive x_1^k, the second x_2^k and so on (of course it only pays if it is captured). We will assume that $x_1^k \geq x_2^k \geq \cdots$.

The agents play in turns, in the same order in each round. Agent α_n is the n-th to play. We chose a consistent (rather than random) playing order so that all the uncertainty in the game was controlled at one point, the new targets. This is somewhat less realistic, but it made it much easier to design and evaluate agent strategies. In some cases, we have commented on how the game would have differed if playing order had been random.

The following are common knowledge:

– The payoffs offered, number of agents required and decommitment penalties demanded by each of the "old" targets.
– The number of new targets that will appear and the number of agents required to capture each.
– The number of agents.
– That agents are perfect reasoners, each one aiming to maximise its own payoff.
– The probability distributions of the payoffs offered by each new target (common knowledge at the start of round one).
– The payoffs offered by each "new" target (common knowledge at the start of round two).

Each simulation consists of a number of runs, each of which has the same first-round targets and the same agents. A new set of second-round targets is generated in each run, according to a fixed probability distribution. We recorded the average average value of the targets captured per run.

Agents use two different brownie-point-like numbers in their calculations. The first one (which we will refer to as "brownie points") is almost exactly like that in [4] — each agent keeps a record of its own behaviour and is uninclined to decommit if its "brownie points" will suffer much. The second *(reliability)* involves keeping a record of other agents' behaviour, and preferring to work with agents who have been reliable in the past. Each agent α updates its record of "Brownie points" and reliability at the end of each run, according to the following rules:

$$\text{BP} = \begin{cases} \text{BP} + 1 & \text{If agent stayed committed} \\ \text{BP} - o_o/o_n & \text{If agent decommitted from a} \\ & \text{target offering } o_o \text{ to one} \\ & \text{offering } o_n \end{cases}$$

$$\text{Rel}(\alpha_i) = \begin{cases} \text{Rel}(\alpha_i) + 1 & \text{If agent } \alpha_i \text{ was at the} \\ & \text{same target and stayed} \\ & \text{committed} \\ \text{Rel}(\alpha_i) - o_o/o_n & \text{If } \alpha_i \text{ decommitted} \\ & \text{from a target where } \alpha \\ & \text{was offered } o_o \text{ and} \\ & \text{moved to a target where} \\ & \text{it was offered } o_n. \end{cases}$$

where BP is the agent's "brownie points", while $\text{Rel}(\alpha_i)$ is its evaluation of agent α_i's reliability. The justification for deducting the ratio of offers from reliability when an agent decommits is that the agent who has been decommitted against is more likely to be forgiving if it didn't lose very much due to the decommitment or if the decommitting agent was making a large gain by leaving. Similarly, in the case of BP, agents are more likely to forgive themselves if they decommitted from a low-paying target to a high-paying one.

When deciding how to play, agents use a weighted sum of the three main considerations: expected payoff, effect on their own "brownie points" and the past behaviour of others (reliability). The decommitment penalties are automatically included in the expected payoff.

Estimating the probability that a target will be captured requires some estimations, since agents can't predict exactly what others will do if the others are considering "brownie points" or reliability as well as their expected payoff (see the discussion in section 4).

3 Some analysis of special cases

The most interesting results in this section are the ones that show that the self-interested agents succeed in achieving the optimal payoff for the group (sections 3.2 and 3.3). Section 3.3 shows that this happens even for quite general assumptions about the decommitment penalties. The first result, in the case where agents have to work in pairs but do not have any decommitment penalties, demonstrates that the agents' behaviour maximises the payoff of the second-least paid agent. This is surprising for a group of agents that are self-interested and do not even consider the payoffs of others.

Note about phrasing: When we write "plays in the first (second) place," we mean "becomes the first (second) agent at a target."

3.1 Case 1: Agents working in pairs behave fairly

Simplifying assumptions:

- No targets appear or disappear after the first round. This means that agents will not change targets in the second round because no new information has become available to them. Hence this analysis is about how they should choose their first target correctly.
- There are no decommitment penalties.
- Each target requires exactly two agents to capture it.
- If a target is captured, it pays strictly positive amounts to each agent at it.
- There are at most twice as many agents as there are targets. If there were more than that, all the agents would know that all the targets would be captured, so their strategy would simply be to take the largest available spot.
- There are only finitely many targets.
- Being the first agent to arrive at any target is always better than being the second to arrive at any target, as long as both targets are captured. That is, $\min_k x_1^k > \max_k x_2^k$.

It is not the case that the first few agents can simply choose to be the first agents at lucrative targets, because they need to worry that those targets may not be successfully captured. Let there be m agents.

Since they are all perfect reasoners, every agent that has an opportunity to guarantee a positive payoff will achieve a positive payoff.

Claim. Every agent, except the last one if m is odd, can guarantee a positive payoff.

Proof If, at agent α's turn, there is a target k with one agent already at it, then α can choose to be the second agent there and guarantee itself a payoff $x_2^k > 0$.

If all the targets are either free of agents or occupied by two, and if there is at least one more agent left to play after α, then α can choose to be the first agent at the target k' with k' chosen so that $x_2^{k'}$ is maximised. If it does this, another agent will surely play at the second place on the same target because at least one of the agents yet to play must play in second place (the last agent always does if it can), and an agent playing in second place can maximise its payoff by playing at target k'.

If m is even, one of the two cases above always holds. If m is odd, one of the cases holds for every agent except the last one. Therefore α can always achieve a positive payoff unless it is the last of an odd number of agents. (Box).

Corollary At the end of the round, every agent (except α_m when m is odd) will successfully capture a target, though not necessarily the one described by the algorithm.

A simple optimal strategy Since each agent has only finitely many possible moves, they could each perform full lookahead to determine which targets would be captured, then choose the most lucrative. Therefore, each of the first $\lfloor \frac{m}{2} \rfloor$ agents to play can work out which targets will eventually be captured, depending on what move it makes. They will always play in the first place at some target, since $\min_k x_1^k > \max_k x_2^k$.

Consider agent $\alpha_{\lfloor \frac{m}{2} \rfloor + 1}$. Playing in the first place at any target must result in a zero payoff, because if a second agent cooperated with $\alpha_{\lfloor \frac{m}{2} \rfloor + 1}$ then one of the preceding agents would end up with a zero payoff, violating the corollary. Even if $\alpha_{\lfloor \frac{m}{2} \rfloor + 1}$ plays as the first agent at a vacant target k such that x_2^k is maximised, no other agent will cooperate — all of the $m - \lfloor \frac{m}{2} \rfloor - 1$ agents that play afterwards will prefer to play in second place at one of the first $\lfloor \frac{m}{2} \rfloor$ established targets. Therefore, the first $\lfloor \frac{m}{2} \rfloor$ agents have chosen (in first place) all the targets offering the best $m - \lfloor \frac{m}{2} \rfloor - 1$ second places.

If m is odd, $m - \lfloor \frac{m}{2} \rfloor - 1 = \lfloor \frac{m}{2} \rfloor$, so the first $\lfloor \frac{m}{2} \rfloor$ agents occupied exactly those targets that had the $\lfloor \frac{m}{2} \rfloor$ highest values of x_2^k.

Therefore the best algorithm for odd m (supposing this is the jth agent to play) is:

If ($j \geq \lfloor \frac{m}{2} \rfloor + 1$) play in second place at an already-occupied target k such that x_2^k is the highest available.

else sort the targets by their values of x_2^k. Choose a target that has one of the highest $\lfloor \frac{m}{2} \rfloor$ values of x_2^k, such that the corresponding value of x_1^k is the highest available.

Theorem 1 *The result always maximises the payoff of the second-last (and hence second-lowest-paid) agent.*

The $m - 1$-th agent occupies the $\lfloor \frac{m}{2} \rfloor$-th best x_2^k position. (Box)

The case when there are an even number of targets is very similar, except that the last agent always has only one possible move, so one first-playing agent can get away with choosing a target that has a small reward for the second agent.

If second-round playing order is random, new information (the playing order) is revealed. If m is odd of and the agents have executed the algorithm above, no agent will benefit from moving in round two because the lucrative places will already be taken. If there is an even number of agents, this is not the case — an agent which played in second place can benefit by moving to the first place at another target and relying on the last agent having only one possible move.

3.2 Case 2: Uncommitted agents working alone behave optimally

Assumptions:

- There are no decommitment penalties.
- Each target requires exactly one agent to capture it.
- If a target is captured, it pays a strictly positive amount to capturing agent.
- There are at least as many first-round targets as there are agents.

Let the values of the "old" targets be $x_1^1, \cdots, x_1^{n'}$ where $n' \geq n$. Without loss of generality, $x_1^1 \geq x_1^2 \geq \cdots \geq x_1^{n'}$. Let the values of the new targets (known probabilistically) be represented by the random variables $X_{(1)} \cdots X_{(n'')}$, where $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n'')}$. The random variable $X_{(i)}$ is the i-th order statistic of the "new" targets' values. Suppose it is given by p.d.f. p_i.

Agents have nothing to lose by committing in the first round, since there are no decommitment penalties. If an agent (apart from the last one) does not commit it may reduce its payoff because another agent playing later may "take its place". Only if agent α_k is certain that at least one new target will pay more than x_1^k does it have nothing to lose by not committing. We will assume such agents commit. All the others will commit because they have nothing to lose by doing so, but the potential to lose if they fail to.

Each agent will chose the most lucrative target in the first round. When new targets appear, the agents will shift into any more lucrative positions in the order they first played. An "old" target that is vacated may be re-occupied by a later agent. Therefore the agents always capture the n most lucrative targets, with their payoffs decreasing in the order of their playing turns.

Agent α_i captures "old" target $i - k$ iff there are exactly k "new" targets more lucrative than it. That is, if $X_{(k)} > x_1^{i-k} > X_{(k+1)}$. The agent captures "new" target $i - l$ iff there are exactly l "old" targets greater than it. That is, if $x_1^l \geq X_{(i-l)} \geq x_1^{l+1}$.

Therefore, at the start of round one, agent α_i's expected payoff is

$$\sum_{k=0}^{i-1} x_1^{i-k} \Pr\left(X_{(k)} > x_1^{i-k} > X_{(k+1)}\right) + \sum_{l=0}^{i-1} \int_{x_1^{l+1}}^{x_1^l} x p_{i-l}(x) dx$$

where $X_{(0)} = x_1^0 = \infty$ and p_{i-l} is the p.d.f. of the $(i - l)$-th best "new" target.

If second-round playing order is random, the analysis is similar except that each agent has to consider its expected payoff in each of the $n!$ equi-probable permutations of players. This makes computation much more intensive and also considerably increases the uncertainty of the agent's payoff. The same observation is true for the following case.

3.3 Case 3: Moderately-committed agents working alone still behave optimally

Assumptions: The same as case 2, except that now there are decommitment penalties. The assumptions about them are:

- Decommitment penalties are always non-negative.

– The decommitment penalty for a target is a function ϕ of the target price, with the property that $id + \phi$ is monotonic increasing.
– If a target is decommitted from, it offers a higher payoff in the second round. Target τ_i offers $x_1^i + \phi(x_1^i)$.

The second condition asserts that the ordering of the targets is the same after adding on the decommitment penalties. This will be true whenever the offers are far apart and the decommitment penalties are relatively small.

If all the agents certainly committed in the first round, the second-round analysis would be the same as case 2 with $x_1^j + \phi(x_1^j)$ substituted for x_1^j for all j. However, since the agents now have something to lose by decommitting, they might not always commit in the first round.

It is possible to arrange a decommitment scheme so that agents are too reluctant to decommit, even when decommiting would be extremely beneficial. This would defeat the purpose of leveled commitment contracts. In this case we can show that such a problem does not occur.

Theorem 2 *In this case (where there are decommitment penalties) it is still true that the n agents occupy the n most lucrative targets at the end of the second round (where the value of "old" target τ_i is taken to be $x_1^i + \phi(x_1^i)$).*

Proof (By contradiction). Suppose there is an agent α_k which, at the end of round 2, occupies a target τ_i that is not among the n most lucrative ones. Then there is at least one unoccupied target in the n most lucrative ones. Therefore when α_k took its turn in the second round, there was at least one unoccupied target τ_j in the n most lucrative ones — the number of unoccupied targets in that set can't have increased since no agents would have left those targets to move to lower-paying ones. If α_k chose τ_i rather than τ_j in round 2, it must have been because

$$O_2(\tau_i) > O_2(\tau_j) \qquad \text{if } \alpha_k \text{ was not already committed to } \tau_i$$

or

$$x_1^i > O_2(\tau_j) - \phi(x_1^i) \qquad \text{if } \alpha_k \text{ was already committed to } \tau_i$$

where $O_2(\tau_k)$ is target k's second-round offer.

These (equivalent) conditions both violate the assumption that τ_j is one of the n most lucrative while τ_i is not (because of the assumption that $id + \phi$ is monotonic increasing. (Box).

In the last two cases, the targets collected when there are decommitment penalties are exactly the targets collected when there are no penalties. Hence decommitment penalties neither improve nor detract from the group's payoff in this case. It is also interesting to ask whether adding decommitment penalties alters which agents capture which targets. In the one- or two-agent cases, there is no change.

Theorem 3 *If there are two or fewer agents, the final state in case 3 (where there are non-zero decommitment penalties) is the same as that in case 2 (where there are none).*

Proof In the one-agent case, the best strategy is to wait until the second round and then pick the best target. This produces the same final state as case 2.

If there are exactly two agents, suppose there are two old targets τ_1 and τ_2 and two new targets σ_1 and σ_2. Assuming two new targets does not result in a loss of generality because if there were more targets, all but the first two would be ignored by both agents and if there were fewer targets, this could be modelled by setting the probability distributions of the two so that one or both of them were guaranteed to be so low that they would not be considered. Assuming two old targets is also not a loss of generality, because we can assign one or both of the targets a value of zero.

Since the final state can only be different from case 2 when in the first round α_2 commits to target 1 while α_1 either doesn't commit or commits to target 2, it suffices to show that it is never individually rational for the agents to behave in this way.

In the following table, "NC" indicates no commitment, "C1" indicates committing to target τ_1 and "C2" to target τ_2 in the first round. $X_{(1)}$ and $X_{(2)}$ (where $X_{(1)} > X_{(2)}$) are the payoffs of the "new" targets, given by p.d.fs p_1 and p_2 respectively. The agents' expected payoffs for all possible round 1 actions are:

It can be seen from Table 1 that α_2's expected payoff depends only on its action, not on that of α_1, although of course α_1's commitments restrict the possible places that α_2 can choose to commit. However, α_1's payoff is dependent on the behaviour of α_2, since it matters whether α_2 commits to target τ_1 possibly forcing α_1 to accept a lower payoff if the "new" targets turn out to be of lower value than expected.

To prove the theorem, we will show that cases 2 and 7 never occur, because α_1 can always predict α_2's behaviour, and if it knows that α_2 will choose to commit to target τ_1 then α_1 will commit to that target instead.

Suppose that α_1 doesn't commit to target τ_1. Agent α_2 only needs to compare committing to τ_1 against not committing, since committing to τ_2 (cases 3 and 5) is always worse than not committing (cases 1, 4 and 6). It will commit to τ_1 iff

$$\begin{aligned}
&\int_{x_1^1+\phi(x_1^1)}^{\infty} p_2(x)[x-\phi(x_1^1)]dx \\
&\quad + \Pr\left(X_{(2)} \leq x_1^1 + \phi(x_1^1)\right)x_1^1 \\
&> \int_{x_1^1+\phi(x_1^1)}^{\infty} x p_2(x) dx \\
&\quad + \Pr\left(X_{(1)} > x_1^1 + \phi(x_1^1) \geq X_{(2)}\right)(x_1^1 + \phi(x_1^1)) \\
&\quad + \int_{x_1^2+\phi(x_1^2)}^{x_1^1+\phi(x_1^1)} x p_1(x) dx \\
&\quad + \Pr\left(X_{(1)} \leq x_1^2 + \phi(x_1^2)\right)(x_1^2 + \phi(x_1^2))
\end{aligned}$$

.

This implies

$$\int_{x_1^1+\phi(x_1^1)}^{\infty} [x-\phi(x_1^1)]p_1(x)dx$$

	Payoffs	
1	α_1(NC):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} xp_1(x)dx$
		$+\Pr(X_{(1)} \leq x_1^1+\phi(x_1^1))(x_1^1+\phi(x_1^1))$
	α_2(NC):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} xp_2(x)dx$
		$+\Pr(X_{(1)} > x_1^1+\phi(x_1^1) \geq X_{(2)})(x_1^1+\phi(x_1^1))$
		$+\int_{x_1^2+\phi(x_1^2)}^{x_1^1+\phi(x_1^1)} xp_1(x)dx$
		$+\Pr(X_{(1)} \leq x_1^2+\phi(x_1^2))(x_1^2+\phi(x_1^2))$
2	α_1(NC):	$\int_{x_1^2+\phi(x_1^2)}^{\infty} xp_1(x)dx$
		$+\Pr(X_{(1)} \leq x_1^2+\phi(x_1^2))(x_1^2+\phi(x_1^2))$
	α_2(C1):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} p_2(x)[x-\phi(x_1^1)]dx$
		$+\Pr(X_{(2)} \leq x_1^1+\phi(x_1^1))(x_1^1)$
3	α_1(NC):	Same as state 1
	α_2(C2):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} p_2(x)[x-\phi(x_1^2)]dx$
		$P \times (x_1^1+\phi(x_1^1)-\phi(x_1^2))$
		$+\int_{x_1^2+\phi(x_1^2)}^{x_1^1+\phi(x_1^1)} p_1(x)[x-\phi(x_1^2)]dx$
		$+\Pr(X_{(1)} \leq x_1^2+\phi(x_1^2))(x_1^2)$
	where	$P = \Pr(X_{(1)} > x_1^1+\phi(x_1^1) \geq X_{(2)})$
4	α_1(C1):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} p_1(x)[x-\phi(x_1^1)]dx$
		$+\Pr(X_{(1)} \leq x_1^1+\phi(x_1^1))(x_1^1)$
	α_2(NC):	Same as states 1 and 6
5	α_1(C1):	Same as state 4
	α_2(C2):	Same as state 3
6	α_1(C2):	$\int_{x_1^1+\phi(x_1^1)}^{\infty} p_1(x)[x-\phi(x_1^2)]dx$
		$+\Pr(X_{(1)} \leq x_1^1+\phi(x_1^1))(x_1^1-\phi(x_1^2))$
	α_2(NC):	Same as states 4 and 1
7	α_1(C2):	$\int_{x_1^2+\phi(x_1^2)}^{\infty} p_1(x)[x-\phi(x_1^2)]dx$
		$+\Pr(X_{(1)} \leq x_1^2+\phi(x_1^2))(x_1^2)$
		(which is always worse than state 2).
	α_2(C1):	Same as state 2

Table 1. Agents' payoffs for the two-player game

$$
\begin{aligned}
&+ \Pr\left(X_{(1)} \leq x_1^1 + \phi(x_1^1)\right)x_1^1 \\
> &\int_{x_1^2+\phi(x_1^2)}^{\infty} x p_1(x) dx \\
&+ \Pr\left(X_{(1)} \leq x_1^2 + \phi(x_1^2)\right)(x_1^2 + \phi(x_1^2))
\end{aligned}
$$

which is exactly the requirement for α_1 to choose case 4 or 5 rather than case 2. That is, it will always commit to target τ_1 when it thinks that α_2 will do so otherwise. Therefore α_1 always captures the most lucrative target and α_2 the second most lucrative. (Box).

4 Empirical Work

When agents are considering "brownie points" or reliability, their behaviour becomes difficult to predict, both for us and for the other agents in the simulation. Since it was no longer feasible to do analysis, we investigated it using empirical methods.

Because the agents cannot predict exactly the behaviour of others, they need to make some estimates of probabilities such as the probability that a particular target will be captured or that a better offer will appear. These estimates are based partly on discrete order statistics [5] and partly on less rigorous methods[1].

Our first experiments were designed to check that our system behaved reasonably when presented with situations similar to those in [1, 2, 4]. These experiments generally replicated the results reported in those papers, showing that a moderate level of decommitment inhibiting (that is a moderate "brownie point" weight in the former, and small-to-moderate decommitment penalties in the latter) were optimal. In both cases, they were better than allowing agents to decommit freely or enforcing full commitment.

The first experiment reported in detail here demonstrates an interesting property of the reliability considerations. It shows that this method can be used to form stable teams that can be more effective overall than ad hoc alliances. In the next experiment, "brownie points" and decommitment penalties are compared on the same scenarios. "Brownie points" are shown to be more effective in both cases, mainly because high decommitment penalties tend to discourage agents from comitting initially. The third experiment shows that having both decommitment penalties and "brownie points" is not better than having only one. The final experiment shows that, although reliability can often be very effective, decommitment penalties are sometimes better.

4.1 Background Experiments: Checking that "brownie points" and decommitment penalties work as expected

Figure 1 shows the agents' average effectiveness in a particular scenario as a function of the weight they attach to "brownie points". There were no decommitment penalites

[1] The main idea is to assume that all the other agents who are not committed will choose targets at random, then to count the various ways in which this can be done. The assumption of random choice is the non-rigorous part, but it seems a reasonable model for behaviour that the agent can't predict

in this experiment. As expected, a medium value (about 40) is the most effective, since then agents tend to decommit exactly when it is particularly beneficial to do so.

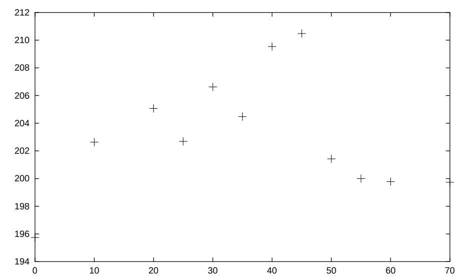

Fig. 1. Average group income vs. “brownie point” weight

Figure 2 shows the results for agents paying no attention to “brownie points”, but having to pay decommitment penalties. All decommitment penalties were equal (since all first-round offers were approximately equal). As expected, a small but non-zero decommitment penalty (about 30) was optimal. This corresponds to 15% of the target’s offered value, which is slightly higher than the percentage noted elsewhere [1, 2].

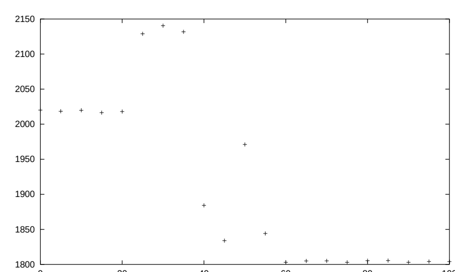

Fig. 2. Average group income vs. decommitment penalty

4.2 Experiment 1: Using reliability to form stable teams

In this case there were no targets in the first round (so an approach based on decommitment penalties would have had no effect), then a large number of equally sized targets in the second round. The number of agents was far too small to capture all the targets, so cooperation was extremely important. Each agent kept a record of every other agent's past decommitting behaviour, summarised by a reliability value that was increased when the other agent successfully captured a target with the first, and decreased when it decommitted against it.

A graph of the agents' average effectiveness over time is shown in Figure 3, for a fixed value of brownie point weighting. Agents were generally ineffective in the first few rounds because later agents ignored low-paying places on targets that were already partially occupied, preferring instead to take higher-paying places on targets that did not yet have other agents. The result was that at the end of a round, there were many targets with not quite enough agents to capture them. As the simulation progressed, those agents who had cooperated successfully in the past (i.e. caught a target together) gradually coalesced into stable teams that always succeed in capturing a target. The teams were much more effective than the disorganised agents, though their behaviour was still below optimal because they often did not capture the best target, just the one that offered the first agent the highest payoff. For reference, the expected optimal payoff (if the agents had some central planner to choose their targets for them) would be approximately 34, about double what the agents achieved.

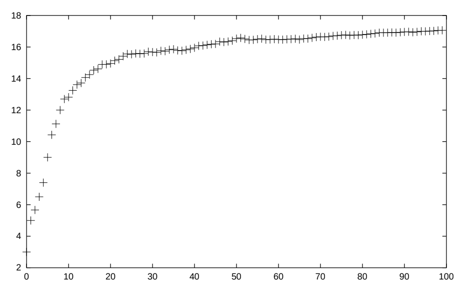

Fig. 3. Running average group income over time for fixed reliability weight

The exact value of the weight attached to reliability seems unimportant, as long as it is non-zero. See Figure 4 for a graph of effectiveness as a function of this weight. The intuition is that after a short while an agent's commitment to its small team becomes an overriding concern, so its exact weight doesn't matter greatly. This is actually a bad thing — if the agents were a little less committed to their team they would be better able to take advantage of the best targets, though they would be less certain of successfully capturing them.

A random playing order would probably disrupt this team formation, because it relies upon agents just happening to appear in the same teams repeatedly, early on in the simulation. It would be interesting to investigate whether teams still formed under random playing order — they would certainly take much longer to do so.

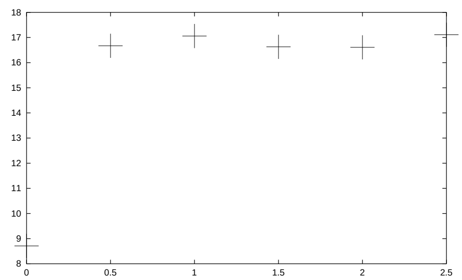

Fig. 4. Average group income vs. reliability weight

4.3 Experiments 2: Comparisons of the different methods on the same scenarios

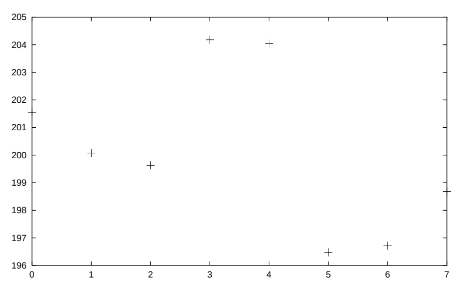

Fig. 5. Average group income vs. decommitment penalty, for Experiment 1 scenario

Figure 5 shows the effect of using decommitment penalties rather than “brownie points” on the same scenario as Experiment 1. The main difference is that the greatest value attained by decommitment penalties is lower than that achieved using brownie points (about 205, compared to 210 in Figure 1). Figure 6 shows the effectiveness of “brownie points” for the scenario in Experiment 2. Again the “brownie points” experiment has a higher average group payoff (c.f. Figure 2).

In the former, where there was one large group target and many smaller “new” ones, decommitment penalties were unsuccessful because they tended to deter some agents from decommitting in the first place. (This was clear from watching the simulation). Agents considering committing in the first round factored the expected loss due to decommitment penalty into their calculations. If the decommitment penalty was high and the agent was likely to get a better offer from the “new” targets, it did not commit. The result was that the group target was rarely captured, so the average group payoff was

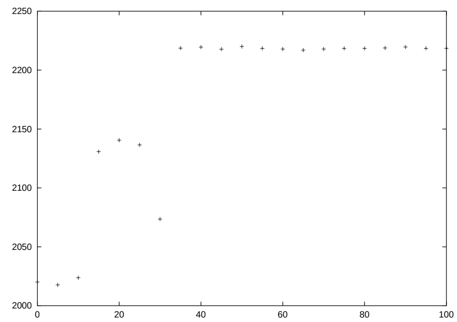

Fig. 6. Average group income vs. “brownie point” weight, for Experiment 2 scenario

not much higher than the payoff of an always-decommitting group. Losing “brownie points” however, was not a disincentive to commit in the first round, just a disincentive to decommit in the second. Hence agents tended to commit to the group target in the first round and then only sometimes decommit to chase the smaller targets. Their group payoff was consequently higher than an always-decommitting group’s.

A similar effect occurred in the latter, though the results were more pronounced because of the scenario itself. Again high decommitment penalties deterred earlier agents from committing in the first round, so the final result was less lucrative than it had been for the “brownie-point” oriented agents. In this case there was a large discrepancy because the usual final arrangement of the agents was different for the different cases — with decommitment penalties, agents tended to capture one “old” and four “new” targets (because of their reluctance to commit to “old” targets initially), but using “brownie points” agents tended to capture three “old” and one “new” target, which was generally more lucrative.

“Brownie points” have the benefits of decommitment penalties, since they gently discourage agents from decommitting, but they don’t have the problem of acting as a deterrent to initial commitment.

4.4 Experiment 3: Combining “brownie points” and decommitment penalties

So far in the experiments described, either there were no decommitment penalties or the weights attached to “brownie points” were always zero. “Brownie points” and decommitment penalties work in very similar ways, by discouraging decommitment just enough that agents decommit only when it is very lucrative. Therefore, intuitively we would expect that combining decommitment penalties and “brownie points” would result in little or no overall improvement, relative to using “brownie points” alone.

This is exactly the result shown in Figure 7, where group effectiveness is plotted against decommitment penalty, for several different “brownie-point” weights. The overall maximum (about 2227 for decommitment penalty 10 and “brownie-point” weight 10) is not significantly higher than the maximum for “brownie points” only (about 2220 for

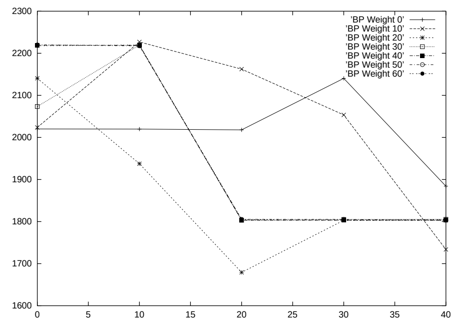

Fig. 7. Average group income vs. decommitment penalty, for different “brownie point” weights

“brownie point” weight 50). As the agents’ “brownie point” weight increases, the optimum decommitment penalty becomes smaller. This reflects the fact that decommitment penalties and “brownie points” combine to limit decommitting.

4.5 Experiment 4: When decommitment penalties work better than reliability

One problem with reliability is that it may take agents some time to form into teams. By contrast, decommitment penalties work immediately. Another issue is that agents using only reliability can sometimes form teams that are too strong, from which they will never decommit no matter how good the outside opportunities. Both of these effects could be seen from the agents’ behaviour in this experiment — Figure 8 shows that agents using reliability were initially much less effective than those using decommitment penalties, and even when they formed into teams were on average slighly less effective.

The scenario consisted of one two-agent target in the first round, with each place offering 3 units, and one one-agent second-round target, offering between 0 and 7 units. Hence the optimum behaviour has expected group payoff $6\frac{1}{8}$, which is about what the decommitment-based group achieved.

5 Conclusions and Further Work

The agents in this scenario displayed some quite complicated and interesting behaviours, including forming stable teams (section 4.2), obtaining optimal group payoff (sections 3.2 and 3.3) and behaving fairly to the less fortunate (section 3.1). The results of the experiments are consistent with other research when the scenarios coincided [1, 2, 4], showing that a moderate “brownie-point” weight or small-to-moderate decommitment penalties produced optimal behaviour.

The most interesting aspect of this work is in further investigations of the team-forming that occurs when reliability is considered. The aims would be to discover a general description for when this happens, and investigate whether it is possible to make the resulting teams more efficient overall.

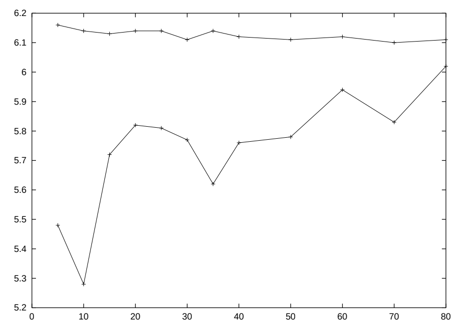

Fig. 8. Average group income vs. number of runs, for agents using only decommitment penalties and agents using only reliability. Those using reliability improve more slowly and reach a lower maximum than those using decommitment penalties.

More generally it would be interesting to try to evaluate more systematically the performance of different methods under many different circumstances, and try to characterise methods that seem to suit particular types of scenario.

The interaction between methods is another direction for research. Intuitively, we would have expected the result in section 4.4 that combining decommitment penalties with consideration of an agent's "brownie points" results in little or no improvement, since the purpose of either method is simply to restrict agent decommitting to those occasions when it is very beneficial. However, adding reliability weights to another method could be very interesting, since reliability has a significantly different effect.

Another interesting question is whether the reliability count is unnecessarily complicated. An alternative approach would be that agents simply remember which agents had decommitted against them and then avoid them completely. This would make the computations significantly simpler, since it would effectively involve a boolean value for reliability. It also coincides more closely with human attitudes to defaulters, at least in some cultures and contexts[2]. It would be interesting to see whether this method was as effective as the more complicated version in causing teams to form.

This scenario could also be considered from the perspective of agent design. So far we have considered only homogeneous agents, but this is not very realistic. One interesting question is what kind of agent is best suited to interact with a given set of (possibly diverse) other agents. Suppose that the designer knows the distribution and types of other agents that will occur. What kind of agent (i.e. what weights for each criterion) should he or she choose? It would be possible to evaluate different types of agents in different contexts using the simulation.

There is still a lot of work to be done before these results can realistically be applied to building intelligent agents for e-commerce. However, the scenario does provide a flexible framework for investigating trust and reliability. Its results so far are consistent with intuition about how a simplified market should behave.

[2] Frank Dignum, personal communication; also [6].

Acknowledgements

This work was completed while both authors were in the Department of Computer Science and Software Engineering at The University of Melbourne. A preliminary version of the work was reported at the Workshop on Artificial Intelligence and Electronic Commerce held in conjunction with the Twelfth Australian Joint Conference on Artificial Intelligence held in Sydney, December 1999.

References

1. Andersson, M. and Sandholm, T. 1998. *Leveled commitment contracting among myopic individually rational agents* in Proceedings of the Third International Conference on Multiagent Systems (ICMAS-98), pp. 26–33, Paris, France
2. Andersson, M. 1998. *Performance of leveled commitment protocols for automated negotiation: an empirical study* Masters Thesis. Royal Institute of Technology, Stockholm.
3. Collins, J. *et al.* 1998. *A market architecture for multi-agent contracting.* in proceedings of the Second International Conference on Autonomous Agents (AGENTS-98), pp 285–292. ACM press, 1998.
4. Glass, A. and Grosz, B. 1999. *Socially Conscious Decision-Making* proceedings of the Bar-Ilan Symposium on Foundations of Artificial Intelligence (BISFAI-99).
5. Kendall, M. and Stuart, A. 1977. *The advanced theory of statistics. Volume 1: distribution theory* Fourth Edition (First edition 1943). Charles Griffin & Co. Ltd. London.
6. K. Kumar, H van Dissel and P Bielli. 1998. *The Merchant of Prato–Revisited: Toward a Third Rationality of Information Systems,* MIS Quarterly, June 1998, pp. 199-226.
7. Maes, P., Guttman, R. and Moukas, A. 1999. *Agents that Buy and Sell: Transforming Commerce as we Know It* Communications of the ACM, March 1999.
8. Nwana, H. *et al.* 1998 *Agent-Mediated Electronic Commerce: Issues, Challenges and some Viewpoints* Second International Conference on Autonomous Agents (AGENTS-98), pp 189–196. ACM press, 1998.
9. Sandholm, T. and Lesser, V. 1996. *Advantages of a leveled commitment contracting protocol* Thirteenth National Conference on Artificial Intelligence (AAAI-96), Portland, OR, pp. 126–133.
10. Sandholm, T., Sikka, S. and Norden, S. 1999. *Algorithms for optimizing leveled commitment contracts* International Joint Conference on Artificial Intelligence (IJCAI), pp. 535–540, Stockholm, Sweden.
11. Sandholm, T. and Zhou, Y. 1999. *Revenue equivalence of leveled commitment contracts* AAAI Workshop on Negotiation: Settling Conflicts and Identifying Opportunities, Orlando, FL, July. AAAI Workshop Technical Report WS-99-12, pp. 38–43.
12. Sen, S. 1996. *Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents,* In Proceedings of the First International Conference on Multiagent Systems (ICMAS-96), pp. 315–321. AAAI Press, 1996.
13. Sen, S. and Biswas, A. 1998. *Effects of misconception on reciprocative agents* in Proceedings of the Second International Conference on Autonomous Agents (AGENTS-98), pp. 430–435, ACM Press, 1998.
14. Steinmetz, E. et al. 1998. *Bid evaluation and selection in the MAGNET automated contracting system.* In Agent Mediated Electronic Commerce, Springer Lecture Notes in Computer Science, volume 1571.
15. Sullivan, D. *et al.*. 1999. *Intention reconciliation in the context of teamwork: an initial empirical investigation* To appear in Proceedings of the 10th International Conference and Workshop on Database and Expert Systems Applications (DEXA-99).