Intention Reconciliation in the Context of Teamwork: An Initial Empirical Investigation

Citation
Sullivan, David G., Alyssa Glass, Barbara J. Grosz, and Sarit Kraus. 1999. Intention reconciliation in the context of teamwork: an initial empirical investigation. In Cooperative Information Agents III, ed. Matthias Klusch, Onn M. Shehory, and Gerhard Weiss, 728-729. Berlin: Springer. Previously published in Lecture Notes in Artificial Intelligence, Special issue, Lecture Notes in Computer Science 1652: 149-162.

Published Version
http://dx.doi.org/10.1007/3-540-48414-0_10

In Klusch, M., Shehory, O., Weiss, G., eds., *Cooperative Information Agents III*, Lecture Notes in Artificial Intelligence, Vol. 1652, Springer-Verlag, 1999, pp. 149-162.

Intention Reconciliation in the Context of Teamwork: An Initial Empirical Investigation*

David G. Sullivan[1], Alyssa Glass[1], Barbara J. Grosz[1], and Sarit Kraus[2]

[1] Harvard University, Cambridge, MA 02138, USA
{sullivan,glass,grosz}@eecs.harvard.edu
[2] Bar Ilan University, Ramat Gan 52900, Israel
and Institute for Advanced Computer Studies,
University of Maryland, College Park, MD 20712, USA
sarit@cs.biu.ac.il

Abstract. With growing opportunities for individually motivated agents to work collaboratively to satisfy shared goals, it becomes increasingly important to design agents that can make intelligent decisions in the context of commitments to group activities. In particular, agents need to be able to reconcile their intentions to do team-related actions with other, conflicting intentions. We present the SPIRE experimental system that allows the process of intention reconciliation in team contexts to be simulated and studied. SPIRE enables us to examine the influence of team norms and environmental factors on team members faced with conflicting intentions, as well as the effectiveness of different intention-reconciliation strategies. We discuss results from pilot experiments that confirm the reasonableness of our model of the problem and illustrate some of the issues involved, and we lay the groundwork for future experiments that will allow us to derive principles for designers of collaboration-capable agents.

1 Introduction

As a result of the ubiquity of computer networks and the phenomenal growth of the Internet, computer systems increasingly are becoming elements of complex, distributed communities in which both people and systems act. Many applications have been proposed that require members of such communities to work collaboratively to satisfy a shared goal (Decker and Li 1998; Sen et al. 1997; Sycara and Zeng 1996). In such situations, agents need to form teams to carry out actions, making commitments to their team's activity and to their individual actions in service of that activity. As rational agents, team members must be able to make individually rational decisions about their commitments and plans. However, they must also be responsible to the team and, dually, able to count on one another. Thus, decision making in the context of teamwork is complex and presents a number of new challenges to the developers of intelligent agents.

This paper focuses specifically on the decision making that self-interested, collaborative agents must perform when their commitment to a group activity conflicts with opportunities to commit to different actions or plans. We describe the initial results of an empirical investigation into the process of intention reconciliation that agents must perform in such situations. The experimental framework we have developed allows us to explore both the effect of team norms and policies on an agent's decisions about

* This research has been supported by National Science Foundation grants IRI-95-25915, IRI-96-18848, and CDA-94-01024. Michael Epstein participated in the development of the SPIRE system. We thank Luke Hunsberger, Martha Pollack, and the reviewer for helpful comments on previous drafts.

conflicting intentions and the effectiveness of various intention-reconciliation strategies that agents can adopt in the face of team norms. Our longer-term goal is to derive principles that system designers can use in constructing computer-based agents that participate in teams. While we recognize that no single approach can adequately meet the needs of every designer in every type of environment, we hope to provide insight into the types of factors that affect individual and team behavior and outcomes, and thus assist developers working in a variety of domains.

2 Intention Reconciliation in the Context of Teamwork

Research on collaboration in multi-agent systems, including our work on SharedPlans (Grosz and Kraus 1996, 1999) and that of others (Levesque et al. 1990; Kinny et al. 1994; Tambe 1997), has established that commitment to the joint activity is a defining characteristic of collaboration. Although theories differ in the ways they encode this commitment, they agree on its centrality. At the same time, research on rationality and resource-bounded reasoning (Doyle 1991; Horty and Pollack 1998; inter alia) has established the need for agents to dynamically adapt their plans to accomodate new opportunities and changes in the environment. However, efforts in this area have mainly focused on plan management and evolution in the context of individual plans. Our work brings these two threads of research together; it addresses the need for collaborative agents to manage plans and intentions in multi-agent contexts, reasoning jointly about commitments to individual plans and commitments to group activities.

2.1 The Problem

Our investigation focuses on the problem of intention reconciliation that arises because rational agents cannot adopt conflicting intentions (Bratman 1987; Grosz and Kraus 1996; inter alia). If an agent has adopted an intention to do some action β and is given the opportunity to do another action γ that would in some way preclude its being able to do β, then the agent must decide between doing β and doing γ: it must *reconcile* intentions, deciding whether to maintain its intention to do β or to drop that intention and instead adopt an intention to do γ.

In particular, we are concerned with intention reconciliation in the context of teamwork, i.e., situations in which at least one of the conflicting intentions is related to an agent's commitment to a team plan. Although "defaulting" on a team-related commitment for the sake of another opportunity may at times appear beneficial from a purely individualistic perspective, agents may need to be concerned with their reputations in the community. The extent to which others trust them not to default may influence their long-term good. An agent must consider how defaulting on team-related commitments may impact its ability to collaborate in the future and, more generally, how team-related factors may affect its future expected outcomes.

We assume that each of the agents is self-interested and acts in an individually rational manner. Even when participating in a collaborative activity, an agent will aim to maximize its own outcome. Agents are also assumed to belong to a community of agents who periodically form teams to accomplish shared goals. Different agents in the community may participate on different teams at different times, and teams may vary in both size and duration. Even though a given team may exist only while engaged in a single group activity, agents in the community may have longer-term relationships. An

α_i : system maintenance of home PC

α_g : system maintenance of large group of workstations

β: upgrade operating system

γ : go to lecture by Nobel Prize winner

Individual context. You have a PC at home; you are the only user. You are committed to doing β in the context of doing α_i. A friend offers you a ticket so you can do γ.

Team context. You are a student employee of the systems administration group at your university and a member of the team doing α_g. You are committed to doing β in the context of doing α_g. A friend offers you a ticket so you can do γ.

Fig. 1. Intention-reconciliation scenarios from the systems administration domain, used to illustrate the differences between individual and team contexts

agent may want or need to participate with other agents in future group activities. Depending on the situation, team members may or may not know each other's identities and contributions to the team. In this work, we do not address the coalition formation problem, i.e., the process by which teams are formed. Furthermore, we use the term *team* to refer to a group of agents who have formed the intentions and beliefs required for collaborative activity. The term *group* refers to a collection of agents that may (or may not) be a team.

2.2 Sample Scenarios

To illustrate the problem of intention reconciliation in the context of teamwork, we will consider an example from one of the domains that our empirical system seeks to model: computer systems administration. Figure 1 sketches two scenarios involving tasks from this domain. In both scenarios, an agent has committed to spending a certain period of time upgrading an operating system (activity β). It is then presented with the opportunity to attend a lecture that occurs during that same period of time (activity γ). Thus, the agent must reconcile a prior intention to do β with a potential intention to do γ. In the first scenario, the prior intention is in the context of a purely individual activity; in the second, the intention is in service of a group activity.

In the individual context, the agent weighs the various costs and benefits of sticking with its original intention or dropping it in favor of the new opportunity. If, for instance, the agent can do the upgrade β the next day without having to drop any other commitments, then it will defer β and commit to going to the lecture. If deferring to the next day means the agent will have to give up going to a movie, then it must also decide whether it prefers the lecture to the movie. On the other hand, if doing β at the planned time is critical to some other activity (for instance, producing a tax return that is due that day), then the agent may decline the lecture ticket. In all these deliberations, only the individual's outcome and future schedule matter.

Similar considerations apply in the team context, but there are additional ones as well. Since the agent's involvement with the systems administration group is an ongoing one, it must consider how other team members will view its failure to honor its commitment to do β. The agent needs to consider the costs it may incur as a result of the team's reaction to its defaulting on a team-related task. In addition, the agent must weigh team-related costs (and benefits) with individual factors.

2.3 Social-Commitment Policies

In interacting with one another, and particularly in working together, we assume that agents in the community adopt, either explicitly or implicitly, what we term *social-commitment policies*. These policies govern various aspects of team behavior, including both rewards and penalties for individual acts in the context of group activities. For instance, they may specify such things as the distribution of benefits from a group activity, the penalty structures imposed on agents who default on commitments to a group activity, and what defines a fair distribution of tasks among agents. We could assume that these policies are agreed on by a team when it forms. However, it seems more natural and efficient to require that the community of agents embody these principles, because in computational settings we expect agent designers will build multiple agents that at different times come together to form different teams.

Social-commitment policies differ from the "social laws" used in other multi-agent planning work (Shoham and Tennenholtz 1992). Social laws provide constraints on agents that allow their actions to be coordinated; these laws constrain the ways agents *do actions* so that their activities do not negatively interact. In contrast, social-commitment policies concern *rational choice* and the ways a society can influence an individual's decision making. As a result, social laws are by their nature domain-specific, whereas social-commitment policies affect decision making across domains and tasks.

2.4 Incorporating Social Factors in Decision Making

Social-commitment policies address the tension between what is best for the individual in isolation and what is best for the team. In this paper we assume agents assess outcomes on the basis of utility functions. Although team members may consider group utility, they do not become group-utility maximizers. By stipulating ways in which current decisions affect future utility as well as current utility, social-commitment policies change the way agents evaluate trade-offs. They provide a mechanism for constraining individuals so that the good of the team plays a role in their decision making. Rosenschein and Zlotkin (1994) have presented similar conventions in the context of negotiation between agents.

Social factors can also function in an additional way. If agents get part of their utility from the team, they have a stake in maximizing group utility. A larger group benefit means a larger share for each agent, and thus a larger individual utility. Therefore, when facing a choice, it may be useful for an agent to consider not only this single choice, but also the larger context of similar choices by itself and others. While being a "good guy" may appear suboptimal by itself, everyone's being a good guy when faced with similar choices may lead to optimal outcomes for everyone in the team. The team as a whole will benefit and each individual ultimately gains. For example, in the team-context scenario of Fig. 1, an individual member of the systems administration team might benefit from choosing to go to the lecture. But if everyone in the team made a similar choice, the group utility would suffer severely. Although such effects could occur within a single interaction (for instance, if the whole team defaults to attend the same lecture), more typically they occur over the longer-term (different members of the team default at different times in favor of such "outside" opportunities). The *brownie points model* described by Glass and Grosz (1999) provides one means of incorporating a good-guy factor into decision making. Policies that encourage good guy behavior are, however, susceptible to manipulation; the "free-rider" problem can arise. Although we recognize this aspect of good-guy behavior, we leave treatment of it to future work.

3 Empirical Framework

3.1 Why Simulations Are Needed

The intention-reconcilation problem outlined above does not seem amenable to a single, all-purpose, analytic solution. Large numbers of agents, the potentially varied capabilities of agents, complex task interactions, uncertainty about future interactions, and incomplete information about other agents all complicate the analysis. Various environmental factors such as the number of tasks scheduled concurrently (task density) also affect outcomes for individuals and teams.

We have thus constructed the SPIRE (SharedPlans Intention-Reconcilation Experiments) simulation system to study the ways in which various environmental factors and social-commitment policies can influence individual and team outcomes and to examine the effectiveness of different decision-making strategies in the face of such environmental and team-related factors. SPIRE is general enough to allow us to model agents from a large set of problem domains, including the two systems we have built based on a SharedPlans-based architecture: WebTrader (Hadad and Kraus 1999) and GigAgents (Grosz et al. 1999).

3.2 The Basic SPIRE Framework

In SPIRE, a team of agents ($G_1,\ldots,G_n$) works together on group activities, called GroupTasks, each of which consists of doing a set of tasks (task instances). Each task instance is of one of the types $D_1,\ldots,D_k$ and occurs at one of the times $T_1,\ldots,T_m$. For example, a GroupTask for a systems administration team that includes both people and software agents might consist of a week's work (with the times T_i being the 40 hours of the work week) doing tasks of the types $D_1,\ldots,D_6$ listed in Fig. 2. Some task-types may have only one instance in the week (e.g., D_6: printer maintenance); others may have multiple instances (e.g., D_5: run and maintain backups). We currently assume that each task type can be performed by a single agent. Agents receive income for the tasks they do; this income can be used in determining an agent's current and future expected utility.

A SPIRE simulation consists of a sequence of GroupTasks. Since varying either the group activity or the team members would make it more difficult to identify sources of variation in the outcomes, we currently require that the same GroupTask be

D_1: read and reply to technical questions by e-mail or in person
D_2: upgrade hardware
D_3: restore deleted files from backups
D_4: check system security
D_5: run and maintain backups
D_6: printer maintenance (paper, toner, etc.)

Fig. 2. Examples of task types from the systems administration domain

done repeatedly by the same team. However, the individual tasks within the GroupTask will not necessarily be done by the same agent each time. SPIRE considers a given GroupTask to consist of a set of tasks with time constraints on the tasks and capability requirements for agents doing the tasks. To simplify the description, we will assume that a GroupTask maps to a "weekly task schedule."

In SPIRE, these weekly task schedules are represented as sets of pairs $<task_i, time_i>$, where $task_i$ is to be done at $time_i$. At the start of each week, a central scheduler takes the elements of the weekly task schedule and assigns them to agents to produce a weekly task-schedule assignment (WTSA).[1] Each agent has a set of task capabilities and a set of available times that constrain the scheduler's assignment of tasks. For instance, only some agents (e.g., humans) might be able to check for security breaks, and only others (e.g., software agents) might be able to run the backup program.

After the scheduler has assigned all of the tasks in the weekly task schedule, agents are chosen at random and given the opportunity to do one of a series of "outside offers." Outside offers correspond to actions that an agent might choose to do apart from the GroupTask. Each outside offer conflicts with a task in the WTSA; to accept an offer, an agent must default on one of its assigned tasks. The central question we investigate is how different strategies for reconciling conflicting intentions (given a particular configuration of social-commitment policies and environmental factors) influence both the rates at which agents default and their individual and collective incomes.

The income values of the outside offers are chosen randomly from a distribution with approximately the same shape as the distribution of task values in the WTS, and with a mean value that exceeds the mean value of the WTS tasks; thus agents have an incentive to default. If an agent chooses an outside offer, γ, it defaults on its originally assigned task β. If there is an available replacement agent that is capable of doing β, the task is given to that agent; otherwise, β goes undone.

The team as a whole incurs a cost whenever an agent defaults; this cost is divided equally among the team's members. The cost of a particular default depends on its impact on the team. At a minimum, it equals a baseline value that represents the cost of finding a replacement agent. If no replacement is available, the group cost is increased by an amount proportional to the value of the task.

3.3 Social-Commitment Policy in SPIRE

For the experiments in this paper, SPIRE applied a social-commitment policy in which a portion of each agent's weekly tasks is assigned based on how "responsible" it has been over the course of the simulation. Each agent has a rank that reflects the total number of times it has defaulted, with the impact of past weeks' defaults diminishing over time. The higher an agent's relative rank, the more valuable the tasks it receives. Since there is a greater cost to the team when tasks go undone, an agent's rank is reduced by a larger amount if it defaults when no one can replace it.

---

1. This central scheduler is used only for convenience. In many domains requiring cooperative agents, agents would most likely need to negotiate each week's schedule. Since this negotiation is beyond the scope of the current SPIRE system and we wish to study aspects of team-commitment scenarios that come after the initial schedule is made, we simplified this aspect of the problem.

SPIRE gives each agent an initial rank of 0, and it uses the following formula to update an agent *a*'s rank at the end of week *i*:

$$rank_a(i) = (PDF) \cdot rank_a(i-1) - penalty\_sum_a(i) . \tag{1}$$

where *PDF*, the penalty-discount factor, is a constant in the range (0, 1) that causes the impact of previous weeks' defaults to lessen over time, and *penalty_sum* is the sum of the rank reductions that the agent incurred because of its defaults during week *i*.

The scheduler assigns *N* tasks per agent on the basis of the agents' ranks. If there is more than one agent with the same rank, the scheduler randomly orders the agents in question and cycles through them, giving them tasks one at a time. Any remaining tasks are assigned to agents picked at random. The strength of the social commitment policy can be varied by modifying the value of *N*.

3.4 Decision Making in SPIRE

In deciding whether to default on a task β so as to accept an outside offer γ, an agent determines the utility of each option. In the version of SPIRE used for the experiments in this paper, the utility that an agent receives from doing an action *act* in week *i* depends on two essentially monetary factors: current income (CI), and future expected income (FEI):

$$U(act, i) = CI(act, i) + FEI(act, i) . \tag{2}$$

Current income only considers the income from the task or outside offer in question, as well as the agent's share of the group cost if it defaults. Its value in the default and no-default cases is thus:

$$\begin{aligned} CI(\mathrm{def}(\beta, \gamma), i) &= value(\gamma) - \frac{group\_cost(\beta)}{n} \\ CI(\beta, i) &= value(\beta) \end{aligned} \tag{3}$$

where def(β, γ) represents the action of doing γ having defaulted on β, and n is the size of the team.

The income that an agent will receive in future weeks depends on its relative position in future weeks' rankings, because higher-ranked agents receive higher-valued tasks. We assume that agents do not know the ranks of other agents, nor the total number of defaults in a given week, but only their own relative ranking in both the current and the previous week. Therefore, an agent can only estimate its FEI, which it does by approximating its new position in the agent rankings both if it defaults and if it does not default, and estimating the assignments it would receive in each case (from the tasks assigned based on rank). By comparing the value of these task sets, the agent can approximate the impact that defaulting will have on its income in the following week.

An agent may also extrapolate beyond the following week when making its FEI estimate. Because the single-week estimate described above is inexact and is less likely to reflect reality for weeks that are further away, an uncertainty factor $\delta < 1$ can be used to discount FEI. Under this approach, if *F* is the original estimate of the following week's income, then the discounted estimate for the *k*th week after the current one is $\delta^k F$. The full FEI estimate in week *i* of an *M*-week simulation is thus:

$$
\begin{aligned}
FEI(act, i) &= \delta F(act, i) + \delta^2 F(act, i) + \ldots + \delta^{M-i} F(act, i) \\
&= (\delta + \delta^2 + \ldots + \delta^{M-i}) F(act, i) \\
&= \frac{\delta(1 - \delta^{M-i})}{1 - \delta} F(act, i)\,.
\end{aligned} \tag{4}
$$

Note that the factor $(1 - \delta^{M-i})$ decreases as the simulation progresses, reflecting the fact that an agent has less to lose from defaulting when there are fewer weeks left in the GroupTask.

Since our current experiments do not consider any "good guy" factors, the utilities that an agent receives from defaulting and from not defaulting in week i of the simulation are given by:

$$
\begin{aligned}
U(\mathrm{def}(\beta,\gamma), i) &= CI(\mathrm{def}(\beta,\gamma), i) + FEI(\mathrm{def}(\beta,\gamma), i) \\
U(\beta, i) &= CI(\beta, i) + FEI(\beta, i)\,.
\end{aligned} \tag{5}
$$

Agents default when $U(\mathrm{def}(\beta,\gamma), i) > U(\beta, i)$.

In another paper (Glass and Grosz 1999), we model the possibility of agents being good guys—i.e., being willing to sacrifice short-term personal gain for the group good—by allowing agents to earn "brownie points" (BP) each time they choose not to default, and including an agent's BP level in its utility function.

4 Preliminary Results

In our pilot experiments with SPIRE, we made the simplifying assumptions that all agents are capable of doing all tasks and that all agents are initially available at all times. To maximize the contrast between "socially conscious" and "socially unconcerned" agents, we also made a relatively large number of outside offers and imposed relatively large rank deductions and group costs when agents defaulted. Figure 3 summarizes the settings used for the majority of these experiments; departures from these values are noted in each experiment's description.

52 weeks per simulation run
12 agents
20 task types (values=5,10,...,100)
40 time slots per week
10 tasks per time slot = 400 tasks per week, of randomly chosen types
10 tasks per agent per week assigned based on the agent's rank, the rest assigned randomly
250-350 outside offers per week:
- number & values chosen randomly
- possible values = task values + 95

initial agent ranks = 0
rank deductions:
- if replacement available, deduct 1
- if no replacement available, deduct 5

discount factor on prior deductions = 0.5
group costs from defaulting:
- baseline=(n/n-1)(max_task_value), where n=# agents
- if no replacement, add (4*task_value)

δ weighting factor for FEI = 0.8

Fig. 3. SPIRE settings used for most of the experiments in this paper. Departures from these values are noted in each experiment's description

The results presented below are averages of 30 runs that used the same parameter settings but had different, randomly-chosen starting configurations (the values of tasks in the weekly task schedule, and the number and possible values of the outside offers). In each run, the first ten weeks serve to put the system into a state in which agents have different ranks; these weeks are not included in the statistics SPIRE gathers.

4.1 Varying the Strength of the Social-Commitment Policy

For all of the experiments, we employed the social-commitment policy described in Sect. 3.3, in which agents are ranked and assigned tasks based on how often they have defaulted. In our first set of experiments, we varied the policy's strength by using different values for the number of tasks per agent, N, assigned on the basis of rank.

Results for $N = 0$, 5, 10, 15, and 20 are graphed in Fig. 4. As expected, the average number of defaults per week drops off as the value of N increases (Fig. 4, *left*). The $N = 0$ case (all tasks assigned randomly) is equivalent to having no social-commitment policy at all. Since defaulting has no effect on FEI in this case, agents are effectively "socially unconcerned" and consider only CI when deciding whether to default on a task. Because outside offers are almost always worth more than tasks—even with an agent's share of the group cost factored in—agents default on average over 90% of the time. Clearly, this situation is undesirable from the point of view of the team.

As N increases, the social-commitment policy drastically reduces the average number of defaults. While this result is unsurprising, it verifies that the FEI estimates made by the agents are reasonable, and it provides a concrete demonstration of how a social-commitment policy can affect the decision making of self-interested agents.

The impact of the social-commitment policy on both mean individual income (from tasks and offers) and group income (from tasks only) is shown in the right half of Fig. 4. Incomes are normalized by dividing by the income that would have been earned if the originally assigned tasks had all been completed. Negative income values can occur as a result of the shared group costs incurred when agents default.

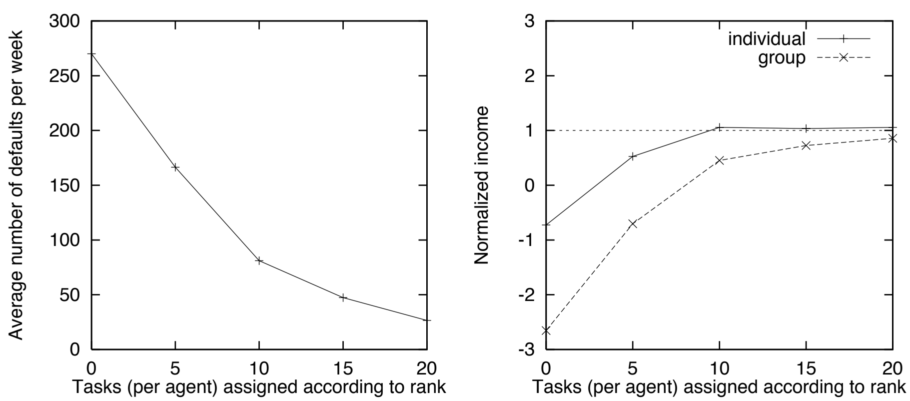

Fig. 4. Effect of the social-commitment policy on the average number of defaults per week (*left*) and on the normalized group income and normalized mean individual income (*right*). Incomes are normalized with respect to the amounts that would have been earned if the originally assigned tasks had all been completed

When all tasks are randomly assigned ($N = 0$), the high number of defaults results in a large loss of group task income, as well as added group costs. Therefore, the group task income is very low (approx −2.7, where 1.0 represents what would have been earned with no defaults). Mean individual income is also negative, but it is higher than group income because of the payments that agents receive for outside offers. This result illustrates that individually rational decisions can still lead to suboptimal outcomes for individuals, in this case as a result of shared group costs. Individuals consider group costs when reconciling their own intentions, but they fail to take into account the costs they will incur from defaults by other agents.

As the value of N increases and agents default less often, both group and individual incomes increase. For $N = 10$, 15, and 20, individual agents do slightly better than they would have if they had done all their assigned tasks. The "plateau" effect that occurs in this range comes from a balance between the value of outside offers and the group costs incurred from defaulting. Agents accept fewer outside offers (and thus lose the extra income that such offers bring), but they also incur lower group costs.

4.2 Varying the Weight Given to FEI

Our next set of experiments varied the δ value that agents use when they weight their single-week FEI estimates (F) to obtain estimates of FEI over the rest of the simulation (cf. Sect. 3.4). As the value of δ increases, so does the value by which F is multiplied, and FEI thus becomes a larger part of the agents' utilities. We therefore expected to see fewer defaults as δ increases. The results shown in the left half of Fig. 5 confirm this. In addition, both mean individual income and group task income again increase as the number of defaults decreases (Fig. 5, *right*).

δ values of 0.4 and 0.5 lead to particularly poor outcomes, since they never multiply the single-week FEI estimate (F) by more than 1, even when there are many weeks left in the simulation. δ values of 0.6, 0.7, and 0.8 are more effective, since for most of

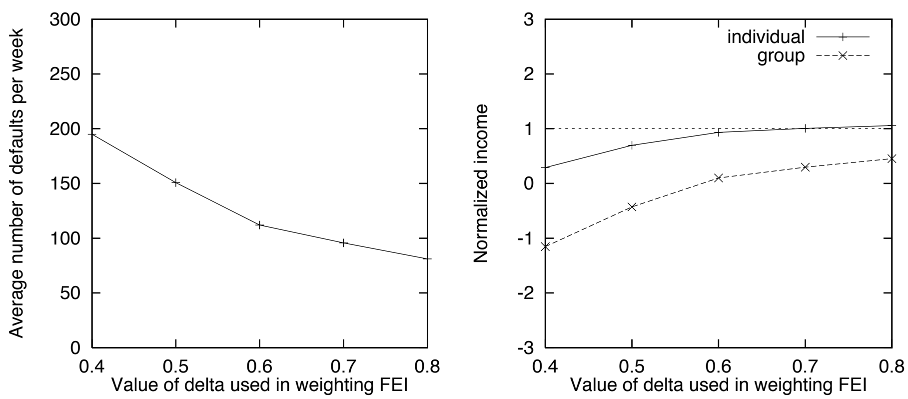

Fig. 5. Effect of the weight given to FEI on the average number of defaults per week (*left*) and on the normalized group income and normalized mean individual income (*right*). Incomes are normalized with respect to the amounts that would have been earned if the originally assigned tasks had all been completed. See Sect. 3.4 for a detailed explanation of the way in which the parameter δ is used

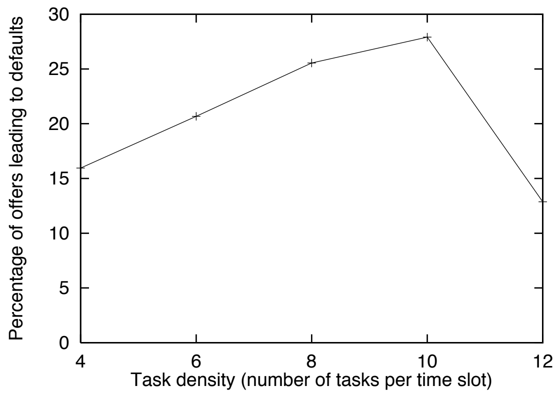

Fig. 6. Effect of task density on the percentage of outside offers that lead to defaults. There were 12 agents throughout, so 12 tasks/time slot is the maximum density

the simulation they multiply F by factors of about 1.5, 2.3, and 4, respectively (see the last line of equation (4)).

4.3 Varying the Task Density

The last of our pilot experiments varied an environmental factor, the number of tasks scheduled in each time slot (task density). Since a larger task density makes it more difficult on average for a defaulting agent to find a replacement, and since the group costs and individual rank penalty are larger when there is no replacement, we expected that there would be fewer defaults as task density increased. However, our results (Fig. 6) do not confirm this hypothesis. Instead, as task density increases, there is a gradual *increase* in the percentage of outside offers for which defaults occurred, with the exception of a drop that occurs at the maximum density of 12 tasks per time slot (with 12 agents). This increase occurs despite the fact that the percentage of offers for which no replacement is available also increases as the task density increases (Table 1).

We were puzzled by these results, until we realized that task density also affects the values of the tasks assigned based on rank, and thus the FEI estimates made by agents. For each of the task densities, we consistently scheduled 10 tasks per agent based on rank (120 tasks in all), and the tasks assigned during this stage of the schedul-

Table 1. Effect of task density on the average percentage of outside offers for which no replacement agent is available

Tasks density	Offers with no replacement
4	0.00%
6	0.02%
8	2.06%
10	39.88%
12	100.00%

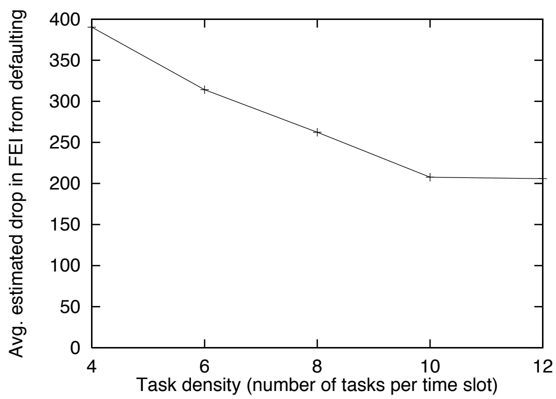

Fig. 7. Effect of task density on the average of the agents' estimated losses in future expected income as a result of defaulting on an assigned task

ing were always the most valuable tasks still available. However, as task density increases, so does the total number of tasks, and thus (on average) the number of tasks of each type. This means that the 120 tasks assigned according to rank tend to have an increasingly narrow range of values as task density increases. As a result, the effect of rank on the tasks an agent receives—and therefore on its FEI—is lessened. In the extreme case, if there were more than 120 tasks with the highest value, an agent's rank would have no effect on the value of the tasks it received.

To confirm this explanation, we analyzed data we collected regarding the agents' estimates of how much FEI they would lose by defaulting. We found that as task density increases, the average estimate of the drop in FEI caused by defaulting *decreases* (Fig. 7), suggesting that the tasks assigned based on rank are indeed drawn from more and more homogeneously valued pools of tasks. In the maximum density case, the fact that replacements are never available makes the average cost of defaulting large enough to outweigh this effect.

This experiment illustrates how a system like SPIRE can uncover unexpected interactions between parameters, enabling agent designers to find them in advance and adjust their designs accordingly.

5 Related Work

Kalenka and Jennings (1999) propose several "socially responsible" decision-making principles and empirically examine their effects in the context of a warehouse loading scenario. Our work differs from theirs in three ways: (1) their policies are domain-dependent and not decision-theoretic; (2) they do not vary environmental factors; and (3) they do not look at conflicting intentions or agents defaulting on their tasks, but at whether agents choose to help each other.

Sen (1996) also considers decision-making strategies that encourage cooperation among self-interested agents, but his work focuses on interactions between pairs of individual agents, rather than those between an individual and a team.

There is also a significant body of economics literature on rational choice and intention reconciliation (Iannaccone 1992; Holländer 1990; inter alia) that space limitations preclude our reviewing here.

6 Conclusions

We have developed an empirical framework that enables us to simulate the process of intention reconciliation in team contexts and to examine the impact of environmental factors and team norms as well as the effectiveness of various decision-making strategies in the face of these external factors. Our initial experiments confirm the reasonableness of our model and illustrate some of the issues involved in the problem we are trying to address.

In a related paper (Glass and Grosz 1999), we investigate agents who consider both their monetary interests and their reputation as team members when reconciling conflicting intentions. In future work, we intend to investigate the following classes of problems within the SPIRE framework: (1) the influence of information about other team members on the agents' behavior; (2) heterogeneous communities, including agents with different capabilities and time availabilities, and agents who embody different decision-making strategies (e.g., some may be good guys, others not); (3) teams with larger numbers of agents; (4) alternative social-commitment policies; (5) alternative intention-reconciliation strategies; and (6) the possibility of agents modeling and adapting to the team behavior of other agents.

Since intention reconciliation in realistic multi-agent contexts is an extremely complex problem, we believe a system like SPIRE is essential for obtaining the insights needed to design collaboration-capable agents (Grosz et al. 1999). Such agents will function not merely as tools but as problem-solving partners, working as members of heterogeneous teams of people and computer-based agents in our increasingly interconnected computing environments.

References

[Bratman 1987] Bratman, M.E. 1987. Intention, Plans, and Practical Reason. Harvard University Press, Cambridge, MA.

[Decker and Li 1998] Decker, K. and Li, J. 1998. Coordinated hospital patient scheduling. In: *Proceedings of ICMAS-98*, pp. 104-111.

[Doyle 1991] Doyle, J. 1991. Rational belief revision. In: *Proceedings of the Second International Conference on Knowledge Representation and Reasoning (KR-91),* pp. 163-174.

[Glass and Grosz 1999] Glass, A. and Grosz, B.J. 1999. Socially conscious decision-making. Submitted to the Bar Ilan Symposium on the Foundations of Artificial Intelligence (BISFAI-99).

[Grosz and Kraus 1996] Grosz, B.J. and Kraus, S. 1996. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269-357.

[Grosz and Kraus 1999] Grosz, B.J. and Kraus, S. 1999. The Evolution of SharedPlans. In Wooldridge, M. and Rao, A., editors, *Foundations and Theories of Rational Agency*. Kluwer Academic Publishers, The Netherlands, pp. 227-262.

[Grosz et al. 1999] Grosz, B.J., Hunsberger, L., and Kraus, S. 1999. Planning and acting together. *AI Magazine* (to appear).

[Hadad and Kraus 1999] Hadad, M. and Kraus, S. 1999. SharedPlans in electronic commerce. In Klusch, M., editor, *Intelligent Information Agents*. Springer Verlag.

[Holländer 1990] Holländer, H. 1990. A social exchange approach to voluntary cooperation. *American Economic Review*, 80(5):1157-1167.

[Horty and Pollack 1998] Horty, J. and Pollack, M.E. 1998. Option evaluation in context. In: *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, pp. 249-262.

[Iannaccone 1992] Iannaccone, L.R. 1992. Sacrifice and stigma: reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, 100(2):271-291.

[Kalenka and Jennings 1999] Kalenka, S. and Jennings, N.R. 1999. Socially responsible decision making by autonomous agents. In: *Proceedings of the Fifth International Colloquium on Cognitive Science*, pp. 153-169.

[Kinny et al. 1994] Kinny, D., Ljungberg, M., Rao, A.S., Sonenberg, E., Tidhar, G., and Werner, E. 1994. Planned team activity. In Castelfranchi, C. and Werner, E., editors, *Artificial Social Systems, Lecture Notes in Artificial Intelligence (LNAI-830)*, pp. 227-256. Springer Verlag.

[Levesque et al. 1990] Levesque, H., Cohen, P. and Nunes, J. 1990. On acting together. In: *Proceedings of AAAI-90*, pp. 94-99.

[Rosenschein and Zlotkin 1994] Rosenschein, J.S. and Zlotkin, G. 1994. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. MIT Press, Cambridge, MA.

[Sen 1996] Sen, S. 1996. Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In: *Proceedings of ICMAS-96*, pp. 322-329.

[Sen et al. 1997] Sen, S., Haynes, T., and Arora, N. 1997. Satisfying user preferences while negotiating meetings. *International Journal on Human-Computer Studies*, 47(3):407-427.

[Shoham and Tennenholtz 1992] Shoham, Y. and Tennenholtz, M. 1992. On the synthesis of useful social laws for artificial agent societies. In: *Proceedings of AAAI-92,* pp. 276-281.

[Sycara and Zeng 1996] Sycara, J. and Zeng, D. 1996. Coordination of multiple intelligent software agents. *International Journal of Intelligent and Cooperative Information Systems*, 5:181-211.

[Tambe 1997] Tambe, M. 1997. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7: 83-124.