Taylor & Francis
Taylor & Francis Group

REGULATION PROBLEMS IN EXPLORER AGENTS

ELPIDA S. TZAFESTAS

Institute of Communication and Computer Systems, National Technical University of Athens, Zographou, Greece

In the present study, we investigate and analyze the behavior of explorer agents. We perform a number of experiments with single and multiple agents and we obtain a number of corresponding results. First, we decouple the agent's functional/motivational system from its cognitive/representational system and show that intricate regulation is necessary to achieve effective and efficient behavior. In the multiple agents case, we extend the single agent behavioral model with a form of sociality selected from a progression of alternatives designed and evaluated. We also show how the general regulation perspective allows for design and analysis of explorer systems. We do not miss to provide variations of the problem and potential applications all along the way.

INTRODUCTION

Exploration is a typical problem encountered in the behavior-based robotics literature and owes its name and formulation to one of the very first projects in behavior-based robotics that aimed—somewhat futuristically—at the exploration of Mars (Angle and Brooks 1990): a set of robotic agents lands on a planet with the mission to explore its surface for samples of minerals having certain properties. The robots arrive in a spaceship that serves as the planetary base in the course of the mission.

Address correspondence to Elpida S. Tzafestas, Institute of Communication and Computer Systems, National Technical University of Athens, Zographou 15773, Greece. E-mail: brensham@softlab.ece.ntua.gr

The mission is accomplished when the whole surface contained within a certain distance from the base is explored. We ought to note that this exploration problem is described as a *sampling* one, where the agents have to pick *some* mineral samples from the sources of interest. However, in practice, in the corresponding literature, the problem is tackled as a *sweeping* one, where the agents have to exhaust the sources of interest (cf. for instance Brooks and Flynn (1989), Beckers et al. (1994)). In both cases, it is a *field coverage* problem: the termination criterion is that *all* the assumed surface should be explored. In the same course of ideas as the researchers in behavior-based robotics, we have adopted the sweeping variant that lends itself to a more worldly instantiation: let us imagine a set of robotic agents thrown in a garage, a subway station, or another delimited area, with the mission to clean away all objects of a given type, usually litter, such as empty soda cans, nylon bags etc. The agents are supposed to return to their base once their mission is accomplished: the subway-cleaner robots are going to wake up and enter in activity outside operation hours, for instance at night, and return to their base definitely once they have cleaned everything, before the reopening of the station. With respect to sampling, sweeping appears therefore a more primitive problem, since it assumes the same functionality for navigation, detection and localization, but without necessitating a sophisticated spatial reasoning: in the case of sampling, the robot has to *remember*, in a way or another, all the sources of mineral that it has already explored, so as to avoid visiting them again, while in the sweeping case, the fact that a mineral source has been visited does not have to be registered in memory as such. The *memorization* or the *marking* of visited positions, explicitly or implicitly with the aid of mechanisms such as traces or pheromones (for example, Steels (1990)), would induce an increase in efficiency and sweeping speed, but it is not part of the sweeping task's description.

The sweeping problem has been tackled until now from a *functional* point of view:

> *Sweeping (1)—Functional point of view*: *How does one or more agents sweep a delimited area to exhaust the sources of interest?*

The answer to this question is a control system, an architecture, that allows an agent to navigate, perceive, and detect mineral etc., in order to sweep the area in question.

A solution such as those encountered in the literature (for instance, Matarić 1992b, 1994) usually draws on social insect behavior modeling (Deneubourg et al. 1990). As such, it comprises a set of fundamental reactive behavioral rules and a random component and even without spatial reasoning or learning, the coverage of the interest field and the exhaustion of the mineral sources are statistically ensured. However, from a more *cognitive* point of view, this functionality alone does not respond to the crucial question:

> **Sweeping (2)—Cognitive point of view**: *How do the agents know that they have swept the whole area, or that they have accomplished their mission?*

In order to answer that question, we have to reformulate the description of the sweeping task, in a way so as to include an expression, analytical or other, that represents the termination criterion, that is the exhaustion of the mineral sources. To this end, it is sufficient to define an environmental or world variable, the density of mineral sources, which characterizes the state of the explored area at any moment. The explorer-sweeper agent's goal becomes therefore to bring the value of that variable to 0. We will see that an agent having a representation of that variable constitutes a simple solution to this description problem.

> **Sweeping—Environmental variable**: *The critical variable that describes the sweeping task is the density of the sources of interest (for example, mineral sources) in the explored area, denoted in what follows as p_w.*

Thirdly, we seek to study the system's operationality, i.e., the relation between the agents' internal architecture and their performance, in order to find/elaborate an architecture that *optimizes* this performance. The operationality criterion that applies to the sweeping task is, quite obviously, the mission's duration: the agents are more efficient if they accomplish their mission earlier.

> **Sweeping—Operational criterion**: *The measure of the agent's performance is the duration of the mission: an agent A is more operational than an agent B, if for the same initial environmental conditions, it accomplishes its mission earlier.*

The (initial) density of the mineral sources in the world thus serves as a *free parameter* of the sweeping task, i.e., an environmental parameter

uncontrollable by the designer and/or the master of the agents. An optimization of the agents' performance would correspond to the minimal mission duration. Moreover, we would like mission duration to be linear on the value of the environmental variable, to allow ourselves to predict and design for performance. The mission's duration can be computed as the sum of the sweeping duration as such (that is, the period of time elapsed up to the point of picking the last remaining sample) plus the additional time necessary for the agent to realize that the whole area is swept and to definitely return to the base. The first term is a function statistically linear on $p_w(0)$, that depends exclusively on the agent's elementary functionality, i.e., on the way it sweeps the area picking up the samples. The *cognitive* system of the agent according to previous definition is responsible for the second term. The two systems, the *functional* (sweeping) and the *cognitive* system, operate in parallel, as we will see in section 2.

> *Sweeping* (*3*)—*Operational point of view*: *What is the behavioral model of a sweeper-agent that* optimizes *the system's performance for all initial environmental conditions? That is, what is the model that allows the agent to terminate its mission as early as possible after the sources' exhaustion, independently of the value of* $p_w(0)$?

In the simulations below, the world under exploration is defined as a square around the central base: the size of the world is therefore the length of the square's edge (unless otherwise stated, the results reported below have been obtained in a 25×25 world). We assume that the base emits regularly an orientation signal, which the agents perceive and use to return to the base. We also assume that the transport capacity of the agents is limited (here it has been set to 30)—this is what forces them to return several times to the base in order to deposit the collected samples. Finally, and since the simulations are discrete in time and space, we assume that every *action* (move, charging of a sample, complete discharge at base) lasts one unit of simulation time.

In the next section we will describe the sweeper agent's motivational system that ensures the fundamental functionality and discuss the tuning of its various parameters. Next, in the section Cognitive/Representational Level: Recursive Satisfaction and Adaptation, we will present the elementary cognitive component that ensures the termination of the mission. After presenting and comparing two adaptation alternatives considered in the section Endogenous and Exogenous Adaptation, we will study in

the section Operational Coupling: Meta-Adaptation and Self-Regulation, the coupling of the agent with its world in an operational manner and we will show the need for a self-regulatory mechanism.

After studying the problem of exploration/sweeping in the single agent case, we will consider the case of multiple explorer agents. The new questions arising are the following: Does the nature of the problem change qualitatively? To what degree does the system's performance improve? Do the agents need some kind of *social action*, an additional mechanism for adaptation, regulation, or other to realize the termination of the task, i.e., the exhaustion of the sources of interest? In all cases, we will not add any additional functionality, such as group transport by multiple agents (Kube and Bonabeau 2000) or traces/markers of field (Steels 1990). As in the case of the solitary explorer, we will neither add any learning mechanisms, instead *the focus of interest of this study is the termination of the sweeping activity.* Figure 1 shows the state of a world under exploration by 10 sweeper agents.

> *Sweeping* (*4*)—*Social point of view*: *Do we need an additional social action when we move from a single to multiple sweeper agents, in order for them to realize collectively the exhaustion of the sources of interest*?

We will study in the section Population Effects the performance of the sweeper agents according to the number of agents. Contrary to what is stated in the relevant literature, it will be concluded that this performance does not show a self-catalytic, thus super-linear improvement, but rather it shows a saturation level and stabilizes itself for large-agent-population sizes. Next, we will study the performance of an agent population having an additional instrumental behavior of spatial dispersion and it will be revealed that not only the dispersion does not improve the overall performance but it makes it degrade due to the instability of the induced spatial configurations. In the section Sociality Models, we will compare several models of *reactive* sociality and we will arrive at the conclusion that the cooperative sociality and the tit-for-tat sociality are the most operational types of sociality for the exploration/sweeping problem. Some strange results obtained with a variant of the tit-for-tat model will also be discussed. In the section Variations of the Problems, two variants of the original exploration problem will be discussed and the article will conclude with a general discussion of the behavioral model and its implications for autonomous agent design.
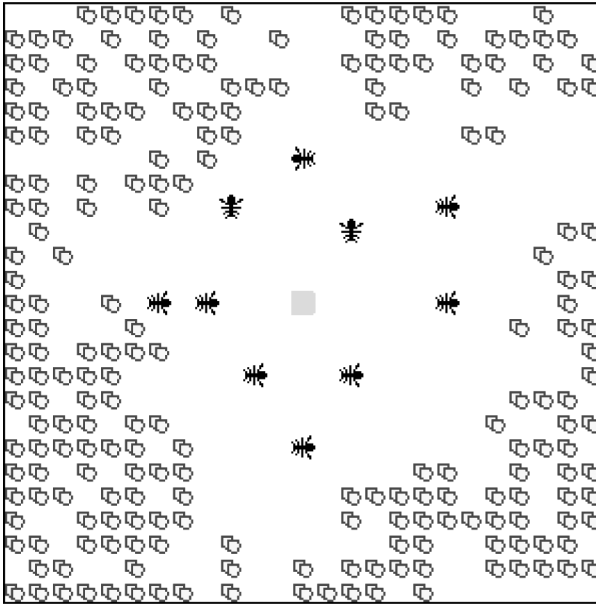
Figure 1. A 25×25 world during exploration by 10 agents ($p_w(0) = 0.9$). The agents are visualized as ants for historical reasons, because the earliest behavior-based models made extensive reference to social insect behavior, especially ants. They pick/consume first the samples that are closer to their base (depicted in the center of the field).

THE SINGLE AGENT CASE

Functional Level: Motivational System

The first implementation of the explorer's functional level has been based on the previous approaches found in the literature, the most prominent example being that of Matarić (1992b) (see also Beckers et al. (1994)). There were two tasks by agent: (1) the foraging (or hunting) task and (2) the return-to-base task. There was also a mechanism of flow or propagation of activation between tasks, inspired by Maes' mechanism (1990, 1991).

This implementation revealed two problems. First, the agent's behavior shows a phenomenon of *an operational cycle* with the trips back and forth to the base, that is *not* explicitly prescribed in the tasks' description, i.e., those phenomenal cycles are emergent. This would not be annoying if our goal were to simply develop the agent's functional system; however, once we turn our interest to the system's operationality, i.e., to its

performance, emergence is no longer enough. What looks necessary is a means to control those otherwise emergent phenomena through a fairly central motivational system.

Second, if the activation flow between the various *rules* (or else the individual actions) that constitute the two tasks is allowed to inflate/deflate in the course of the propagation, the comparison of levels of activation and the competition between tasks loses its balance, and the design of the activation propagation mechanism becomes problematic. This point has been unraveled and analyzed extensively by Tyrrell (1993a, 1993b, 1994) who suggested hierarchical architectures and eventually central action selection systems.

We have thus defined two structural/behavioral units, called tasks, that compete with each other by expressing each an individually computed preference or motivation to execute and use actuator resources. We call the two tasks *homing or return-to-base task* and *hunting or foraging task*, respectively. The homing task commands the robot to head toward the base, if it's not empty, while the hunting task commands the robot to pick an encountered sample or to choose the most promising hunting direction.

The two tasks need to compute locally their motivations that will next be compared and arbitrated in the actuation systems' level, here the navigation system. Those motivations need to depend on a internal variable (or *drive* or tendency to execute) as well as on one or more external stimuli (that correspond to the future expected *reward*). The *goal* of this hedonistic agent is to bring both motivations to 0. According to the ethological literature (cf. for example McFarland and Bösser (1993)), the motivation's computation formula ought to be either additive or multiplicative according to the two measures, the levels of internal drive and the external stimuli:

Motivated Task

$$motivation = drive^* f(stimulus) \tag{1}$$

$$or, \ k^* drive + (1 - k)^* f(stimulus) \tag{2}$$

$$f(stimulus) = (stimulus + a)/(stimulus_{max} + a) \tag{3}$$

To allow balanced competition between the various motivations, we need to normalize all variables (here they are normalized between 0 and 1).

For an additive rule of motivation computation (2), the normalization necessitates another normalized parameter (k) expressing the relative weight of the drive and the stimulus. Furthermore, the normalization of the stimulus ((f *stimulus*), (3)) necessitates an amplification factor (a) expressing the default preference of the agent for that task ($a/(stimulus_{max} + a)$), i.e., expressing the possibility of a *vacuum activity*, or else activity in the absence of stimulus. The parameter $stimulus_{max}$ is an endogenous parameter of the agent, namely the saturation value of the corresponding sensor (the maximal signal value it gives). Note that the real situation is lightly more complicated than what this formula allows to see, because there actually exist *two* sensors, one for the appetitive behavioral component and one for the consummatory component (McFarland and Bösser 1993). It is therefore sufficient to simulate that system by defining as output value of the consummatory component sensor (which is binary by definition) the saturation value of the appetitive sensor.

Note also that the multiplicative motivation computation formula (1) applies to the case of motivations whose drive and stimulus are engaged in an OR relation (motivation is 0 if either drive *or* stimulus is 0), whereas the additive formula (2) applies to the case of motivations whose drive and stimulus are engaged in an AND relation (motivation is 0 if both drive *and* stimulus are 0).

Obviously, the drive of the homing task is the ratio charge/transport capacity (more an agent is charged, more it is motivated to return to its base), while the hunting drive is complementary to the former. On initialization, the agent's charge is 0, so its motivations for homing and hunting are respectively 0 and 1; the agent sets off from the base to hunt and pick mineral samples. The homing task is of type (2), since the agent has to be at the base *and* have a charge 0 in order for the homing motivation to be 0, and cannot show any vacuum activity ($a = 0$). The hunting task is of type (1), since the agent needs a zero drive *or* a zero stimulus for the task's motivation to be 0, otherwise it will resort to vacuum activity (hence $a > 0$). The tweaking of the a and k parameters has been made in a way to prevent the agent from returning to the base prematurely, i.e., when it is only a little charged (cf. detailed analysis in the end of this section). The homing stimulus is the orientation signal emitted by the base and its intensity in a given point in space is the grid size (size of the edge of the grid) minus the perception distance normalized by the grid size. As a consequence, the closer an agent is to the base, more it will be prone to return to it to discharge. On the other hand, if it finds

itself far away from the base, the homing motivation can prevail despite its weak stimulus only if it has a high enough drive, or equivalently if the agent is full or almost full. The hunting stimulus relies on the perception of the mineral sources' density at a distance, in a maximal perceptual hunting range that has been set to three grid cells.

Homing:

$$drive_{homing} = charge/capacity \quad (2)\, k = 0.9, a = 0$$
$$stimulus = (grid\ size - distance)/grid\ size \quad (max = 1)$$

Hunting:

$$drive_{hunting} = 1 - drive_{homing}, \quad (1)\, a = 10$$
$$stimulus = min(density/distance, 10) \quad (max = 10)$$

We can remark that the parameter a expresses the agent's vacuum activity as well as the task's persistence (its dominance beyond the point where it is of absolute priority). Persistence is also reinforced by a noise factor involved in the perception of mineral sources density (with a 5% probability, the perceived density at a distant point is incremented by 1). This persistence is not prescribed in the agent's architecture, it is therefore emergent; Beer and Chiel's (1990) neural architecture demonstrates such an emergent persistence, decaying in time, but unlike ours, the drives are not explicitly represented in the architecture, i.e., there is no motivational system.

A last observation is that the two tasks' drives are coupled in a way that the one's *satisfaction* makes the other rise automatically. So, in the absence of other mechanisms the agent's activity and its round trips to the base will be maintained forever, i.e., the agent's satisfaction state (homing motivation = hunting motivation = 0) is an unreachable state. Inversely, if we want to maintain the agent's activity forever, all that is necessary is to couple its drives in that way. For example, Steels (1995) described in the very same way a parasite elimination task—with the only difference that coupling was situated outside the agent, it did not refer to the internal drives but to the external stimuli. The homing parameter k must be over 0.5 and close to 1, otherwise the agent keeps too close to the base and terminates prematurely. The higher the k value, the higher the task speed. The same holds for the hunting parameter a, which must be close to $stimulus_{max}$, to ensure high task speed (due to
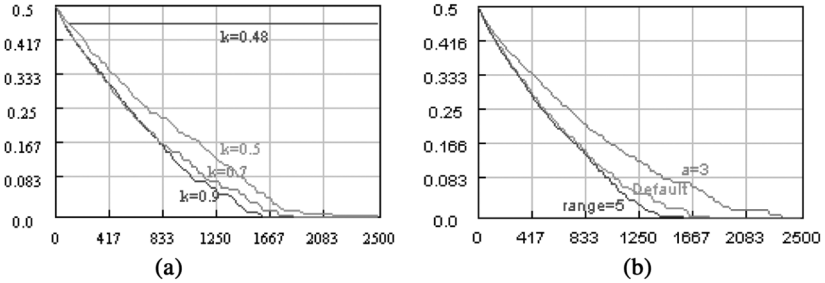
Figure 2. (a) Task speed (p_w curve against time) for k = 0.5, k = 0.7, k = 0.9 ($p_w(0) = 0.5$, $p_a(0) = 0.1$)). (b) Task speed (p_w curve) in three cases: default, a = 3, range = 5 ($p_w(0) = 0.5$).

higher tendency to explore). Also, obviously, the bigger the hunting range, the higher the task speed. These are depicted in Figure 2, where higher task speed shows as steeper slope for p_w. Similar experiments may be performed with different signal or drive formulas.

Cognitive/Representational Level: Recursive Satisfaction and Adaptation

We come now to the second question: How do the agents know they have swept the whole area in order to return to the base? They need a way to detect the degree of task completion or else a termination criterion (sweeping completed). The agent's hunting motivation will then depend recursively on that criterion. Once its value falls to 0, the agent will return definitely to the base, because of the nonzero homing motivation. At this point, both motivations will be 0 and the robot will not move away from the base anymore. Recall that according to that arrangement, the two behavioral components, the picking system and the *cognitive* system are independent and function in parallel, they can therefore be studied and evaluated separately.

Recursive Motivation (as before but multiplied by 'term'):

$$motivation = drive^* f(stimulus)^* term$$

$$or, (k^* drive + (1 - k)^* f(stimulus))^* term$$

$$term = termination\ criterion$$

The only parameter of the task that can be useful to the development of a termination criterion is the source density in the world $p_w(T)$. If the agent knew in advance its initial value $p_w(0)$, we could define as termination criterion a formula such as [$p_w(0)^* d^2$ *samples have been collected*] (where *d* is the size of the square's edge, here 25). However, this criterion is not safe because if a sample is not detected, the agent will never terminate (on the other hand, we could certainly allow ourselves to miss a couple of samples). Another more hybrid criterion is a formula such as [$p_w(0)^* n^* d^2$ *samples have been collected, or* $p_w(0)^* m^* d^2$ *time units have elapsed*]. The first term of the new termination criterion combines the previous formula together with an additional tolerance factor $n \in (0,1)$, while the second term represents a time-out criterion.

The time-out depends obviously on the size of the explored world, so that the *m* factor needs to be sufficiently larger than 1 to allow multiple round trips to the base as well as an occasional erratic and exploratory behavior. The first problem that such a criterion presents is that the value of $p_w(0)$ is not known in advance. The second problem concerns the tweaking of the *m* and *n* parameters. A simple solution to both problems is to estimate continuously the value of $p_w(t)$ and, given that it falls to 0 as a side effect of the agent's activity, take as a termination criterion $p_w(t) = 0 (p_w(t) < \varepsilon)$. Estimation of the value of $p_w(t)$ involves then a representational variable which is local to the agent ($p_a(t)$) and may be done through a simple formula of proportional adaptation using a window of observation *w*:

Representational Variable: $P_a(t)$
Proportional Adaptation (*window w, rate r*):

$$p_a(t) = p_a(t - w) + diff^* r$$
$$diff = p_{comp} - p_a(t - w)$$
$$p_{comp} = number\ of\ picked\ samples/number\ of\ steps\ (moves)$$
$$(during\ the\ adaptation\ window)$$

Termination Criterion:

$$p_a(t) < e_p, where\ e_p\ is\ a\ small\ threshold\ (here, e_p = 0.001)$$

The p_{comp} is the agent's estimate of p_w as computed during the adaptation window and the proportional law ensures that the estimate's update does
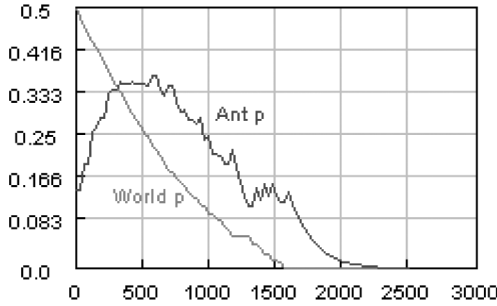
Figure 3. Coevolution agent-world: the agent is back at $t = 2512$ $(p_w(0) = 0.5,$ $p_a(0) = 0.15,$ $w = 30,$ $r = 0.15$). The curve that falls linearly is that of $p_w(t)$. The bell-like curve is that of $p_a(t)$.

not take place too quickly. This adaptive representation system shows the advantage of robustness in front of perturbations/manipulations such as reinitialization of $p_w(t)$ during sweeping, which is unlike what happens with the two previous termination criteria—this is due to the fact that the estimation is a continuous process and does not rely on ad hoc variables, such as a time-out variable. Figure 3 illustrates the co-evolution of the two variables $p_w(t)$ and $p_a(t)$. As is shown in the figure, *the representational variable allows the agent to always solve its termination problem without ever taking the real value of the variable it represents* (except a crossing point). Both variables fall progressively to 0 without ever taking the same value—we could say that $p_a(t)$ *follows* $p_w(t)$. Actually, the rapid rise of $p_a(t)$ in the beginning of the sweeping phase is due to the presence of a sensor of distant samples that makes the agent head toward the mineral sources minimizing its erratic behavior in a way that most of the visited places contain samples (for the same reason the percentage of real field coverage at task termination is typically somewhere between 45% and 75%). The value of $p_a(t)$ falls then because the value of $p_w(t)$ decreases as a side-effect of the agent's activity who finds less and less samples.

Note also that this coupling between agent and world (that shows as a coupling between $p_a(t)$ and $p_w(t)$) with the definition of a motivation and a recursive satisfaction state, may make the agent unable to arrive to its satisfaction state if the world is constantly perturbed (we can maintain the agent's activity forever, if we systematically reinitialize the world every time it is almost empty).

To return to the utilitarian example of the robot that collects litter in a subway station, we can imagine an additional mechanism that would reinitialize the value of $p_a(t)$ at given moments (for instance every morning at 3 am) so as to force the agent to start sweeping. Such a mechanism would thus give rise to a periodical phenomenon that would be perceived as a *small death* of the robot every night.

Endogenous and Exogenous Adaptation

A second adaptation mechanism has been also studied. This mechanism is more *exogenous* than the previous one, in that it depends on a perceived measure, rather than on an activity measure computed inside the agent:

Exogenous Adaptation:

$$p_{comp} = mean\ of\ perceived\ samples/number\ of\ steps\ performed$$
$$(during\ the\ adaptation\ window)$$

The comparative results of the two mechanisms are given in Figure 4. The endogenous adaptation mechanism (adaptation by activity) gives more idiosyncratic and predictive results, whereas the exogenous adaptation mechanism (adaptation by perception) makes the agent more manipulable, since $p_a(t)$ is behind $p_w(t)$. Manipulability of the exogenous adaptation criterion is manifest in that the agent is less robust to
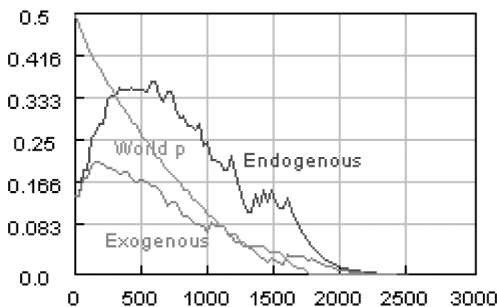

Figure 4. Endogenous adaptation (by activity) versus exogenous adaptation (by perception). $p_w(0) = 0.5$, $p_a(0) = 0.15$. Duration of the two experiments: $t_1 = 2512$, $t_2 = 2421$, $w = 30$, $r = 0.15$. In the second case, the agent is much more manipulable, because it is behind the world (also see text).

short-term perturbations: we have been able to maintain the agent's activity by giving it regular false stimuli (illusions), that is by placing samples in spots where it could perceive them and withdrawing them once the agent had come close.

Operational Coupling: Meta-adaptation and Self-regulation

Next, we proceeded to study the relation between the adaptation parameters (window w and rate r, cf. the section Cognitive/Representative Level: Recursive Satisfaction and Adaptation) and the initial world value $p_w(0)$. The system has been simulated for several values of w and r in several initial world densities. The simulation results for three sets of adaptation parameters (quick, medium or slow adaptation) are given in Figure 5.
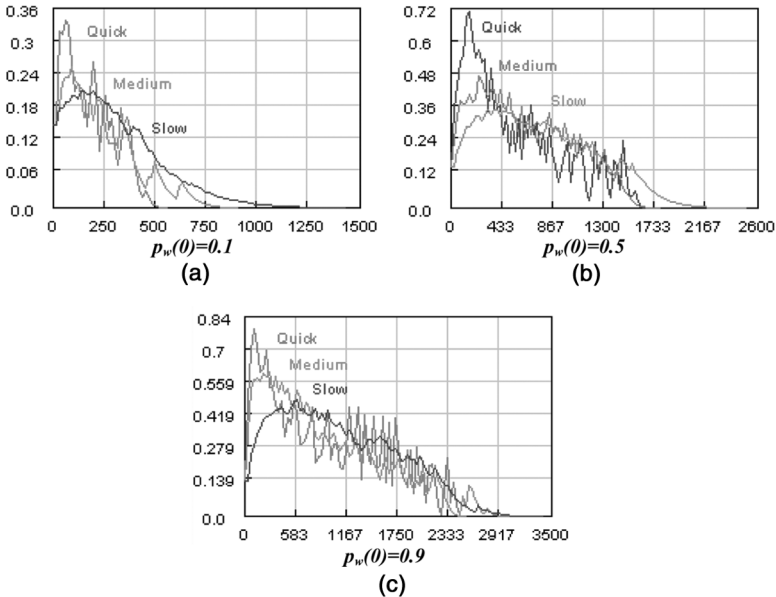

Figure 5. Performance of the agent for three different parameter settings ($p_a(0) = 0.15$). Quick adaptation: $w = 10$, $r = 0.3$. Medium: $w = 15$, $r = 0.2$. Slow: $w = 20$, $r = 0.1$. (a) $p_w(0) = 0.1$, (quick) $t_1 = 562$, (medium) $t_2 = 914$, (slow) $t_3 = 1428$, (b) $p_w(0) = 0.5$, (quick) $t_1 = 1705$, (medium) $t_2 = 1785$, (slow) $t_3 = 2504$, (c) $p_w(0) = 0.9$, (quick) $t_1 = 2540$, (medium) $t_2 = 3017$, (slow) $t_3 = 3404$.

The quick adaptation is more operational than the medium one, which is in turn more operational than the slow one (always according to the first section's criterion). However, the quicker the adaptation, more fluctuations it shows, and the slower the adaptation, more delays it shows. Furthermore, the same parameter setting gives different results in different world densities: the difference in the results is reflected on the shape of the curves. More particularly, the agent's response to different perturbations (the shape of the curve of $p_a(t)$) differs according to the free task parameter ($p_w(0)$): for the same parameter setting, the agent finishes its task more or less quickly according to the value of $p_w(0)$, that is the duration of the interval between the moment of picking of the last sample and the definitive return of the agent to the base is very variable. It seems therefore that to ensure the agent's operationality in different worlds, we need to find a means to combine the operational advantages of quick adaptation with the advantages of slow adaptation as far as curve regularity is concerned. More precisely, we need a quick adaptation near the end (to terminate quickly), but a slow adaptation during picking (to avoid fluctuations). We have then to find a way to stabilize to the right parameter setting *on-line*. The solution is to have a range of possible values for each of the two adaptation parameters and an additional criterion for online value updating, that is for self-regulation within the parameter range limits. Otherwise stated, *we need a meta-adaptation system*.

Meta-adaptation has to affect the w and r parameters in a way that adaptation becomes quicker when p_{comp} is sufficiently close to $p_a(t)$ and slower when it is far from it. This meta-adaptation law translates the fact that the world is more reliable when it is not much different from the agent's idea about it, otherwise the world should not be taken too seriously. Some experiments with the opposite law (quicker adaptation when the world differs much) have proved it to be counter-intuitive and nonoperational, since it would make the agent more manipulable in capricious worlds.

Meta-adaptation:
If $|diff| (= |p_{comp} - p_a(t - w)|) \leq f_P$,
then quicker adaptation
$$r \rightarrow r_{max}, w \rightarrow w_{min}$$
otherwise slower adaptation
$$r \rightarrow r_{min}, w \rightarrow w_{max}$$

(1) *Gradual mode*:

Quicker $r = r + r_r * (r_{max-r})$, $w = w + r_w * (w_{min-w})$

Slower $r = r + r_r * (r_{min-r})$, $w = w + r_w * (w_{max-w})$

(2) *Bang bang mode*:

Quicker $r = r_{max}$, $w = w_{min}$

Slower $r = r_{min}$, $w = w_{max}$

Figure 6 gives the results of applying the gradual meta-adaptation system in three initial world densities; as is shown in the figure, the agent's response (the shape of the curve) is the same for all three exemplary densities, or else the residue of mission duration after picking the last sample is approximately the same in all three cases.

The curves of the adaptation window and rate for the experiment of Figure 6 are given in Figure 7. We observe that the agent becomes more adaptive (its adaptation is quicker) toward the end of sweeping: this is due to the considerable decrease of $p_a(t)$ that approaches 0, so that the meta-adaptation criterion is satisfied and adaptation becomes more and more rapid up to its maximum. For the same reason, the agent remains only a little adaptive during sweeping/picking, which allows it to better follow $p_w(t)$ and to make the curve of $p_a(t)$ look fairly regular, without the significant fluctuations/delays inherent in Figure 5.
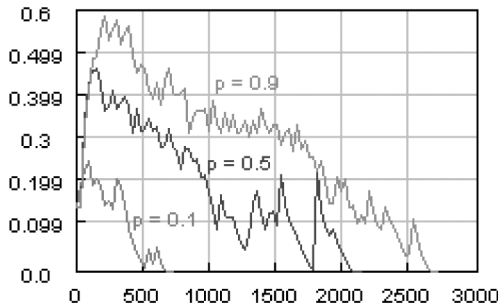

Figure 6. Performance of the agent with a gradual meta-adaptation system for three initial world densities, low $(p_w(0) = 0.1)$, medium $(p_w(0) = 0.5)$ and high $(p_w(0) = 0.9)$ $(p_a(0) = 0.15)$. $t_1 = 739$, $t_2 = 2137$, $t_3 = 2724$. $(f_p = 0.1$, $w_{min} = 10$, $w_{max} = 20$, $r_{min} = 0.1$, $r_{max} = 0.3$, $r_r = r_w = 0.2)$. The shapes of the curves are more irregular than that of Figure 3 because adaptation is on average more quick here (recall that in Figure 3 we had $w = 30$, $r = 0.15$). For instance, the sudden upward rise near the end of $p = 0.5$ curve is due to the discovery of the last isolated samples when the agent was almost finished and the abrupt adaptation of p_a accordingly.
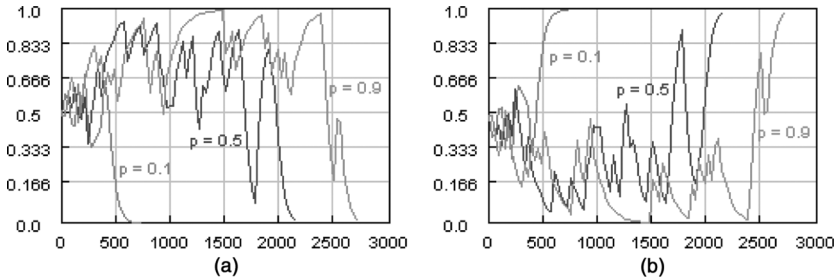
Figure 7. (a) The curve of the agent's adaptation window *w*. Its value is normalized between 0 and 1, where 0 corresponds to the minimal value ($w_{min} = 10$) and 1 to its maximal value ($w_{max} = 20$). (b) The curve of the agent's adaptation rate *r*. Its value is normalized between 0 and 1, where 0 corresponds to the minimal value ($r_{min} = 0.1$) and 1 to its maximal value ($r_{max} = 0.3$).

   We have then examined the case of bang-bang meta regulation, where *r* and *w* assume immediately their *max* or *min* values, as opposed to the previous gradual mode of regulation which assumed convergence toward *max* or *min*. Bang-bang meta regulation (Figure 8) is clearly superior to gradual, since it is more operational in the end (compare termination values in Figures 6 and 8) and it shows less fluctuations in the middle (compare curves in Figures 6 and 8). This observation confirms the intuitive idea that the actual values of the adaptation parameters *w, r* do not count much in themselves, but the decision about quick/slow adaptation is all that is necessary. Therefore, when it makes sense to adopt a slow adaptation scheme, it is not necessary to converge to the slow values, but rather to use them right away, otherwise the system might appear indecisive at times.
   Other experiments have been carried out to explore the relative importance of the adaptation rate and the adaptation window: the system has been simulated in the case where meta-adaptation acts only on the rate or only on the window. The first alternative led to a much more operational system than the second alternative while still a little less operational than the default case.
   The meta-adaptation has subsequently been tested in the case of exogenous adaptation at the first level: the results where analogous to those obtained in the endogenous adaptation case (better operationality and independence from the initial world density), but the disadvantages in terms of manipulability have remained intact.
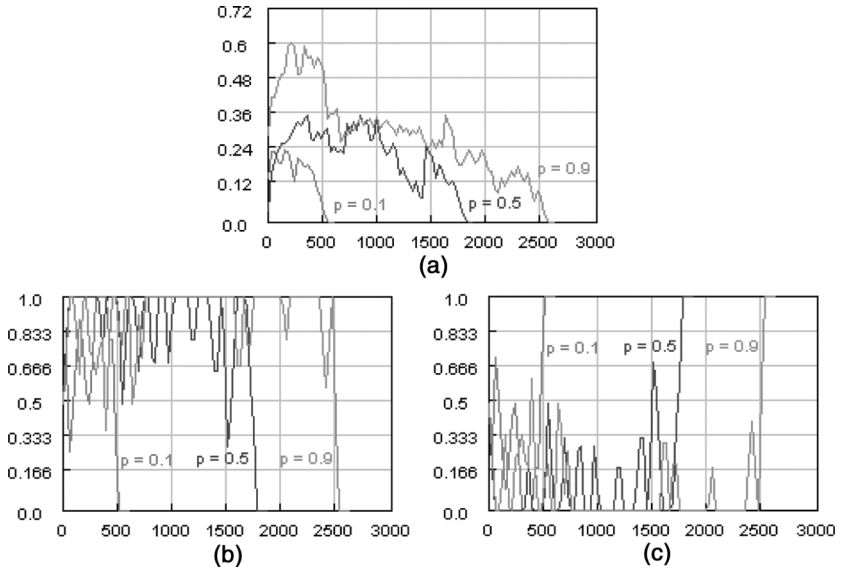
Figure 8. (a) Performance of the agent during the same experiment as in Figures 5 and 6, but with bang-bang meta mode. $t_1 = 627$, $t_2 = 1885$, $t_3 = 2628$. (b) The curve of the agent's adaptation window w. Values are normalized between 0 and 1, where 0 corresponds to minimal value ($w_{min} = 10$) and 1 to maximal values ($w_{max} = 20$). For visualization purposes, the figure shows the moving averages of 10 cycles. (c) The curve of the agent's adaptation rate r. Values are normalized between 0 and 1, where 0 corresponds to minimal value ($r_{min} = 0.1$) and 1 to maximal values ($r_{max} = 0.3$). For visualization purposes, the figure shows the moving averages of 10 cycles.

MULTIPLE AGENTS

Population Effects

We have simulated the system with 1 to 100 agents without any additional *social* functionality or action and we have recorded the task completion time, i.e., the time from start of the task until picking of the last sample. It should be stressed that we have adopted as termination point of the experiment, the moment of picking of the last sample rather than the moment of return of the last agent to the base, in order to eliminate the *cognitive* problem of the previous section (*How do they know they have finished?*): in the absence of other collective functionalities, this point is independent of the cognitive properties of the agents, in other words the sweeping of samples is parallel with the operation of the representational/cognitive
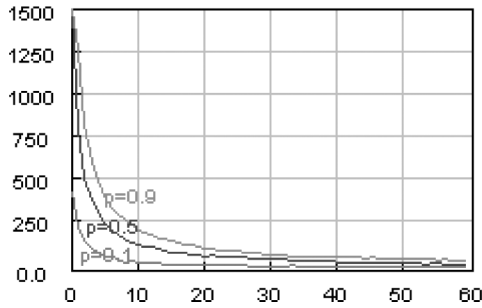
Figure 9. Task completion time according to the number of explorer agents ($p_w(0) = 0.5$) for three initial world densities, low (0.1), medium (0.5), and high (0.9).

system (cf. the Introduction and Section Cognitive/Representational Level: Recursive Satisfaction and Adaptation).

Our purpose has been to show that in the absence of explicit sociality the performance of the system does not improve in a super-linear manner as is said in (Matarić 1992a) and (Arkin et al. 1993). Actually, the dispersion mechanism is not in any sense self-catalytic, consequently the only possible gain is the gain from parallelism, which can be linear at best. Tzafestas (2001) shows that even with self-catalytic mechanisms, the performance improves up to a saturation level: for very large populations, there is significant overcrowding.

Figure 9 shows the performance of the system and how this stabilizes eventually without showing super-linear improvement: the figure gives the results for 1 to 60 agents—performance rises quickly to a saturation value at about 15 agents, while beyond 60 agents it remains stable. We will then seek to accelerate the task execution (sweeping and return to the base) by introducing an additional dispersion functionality or a regulatory cooperation mechanism.

The Role of Dispersion

Matarić (1992a) introduced in her explorer robots a dispersion behavior to accelerate the field coverage and consequently the task execution speed.

Dispersion Rule (*Matarić 1992b, p. 437*)
If a robot is perceived, turn to avoid and advance for a fixed period of time.

If multiple robots are perceived, turn away from their local centroid and advance for a fixed period of time.

Matarić (ibid.) separated the dispersion from the other behaviors of foraging and flocking and used a mechanism of activation propagation to arbitrate them; however, she wrote herself (Matarić 1994, p. 47) that this configuration is perhaps not minimal.

During our first experiments (cf. see section Functional Level: Motivational System) we sought a generic dispersion behavior, that would yield an optimal spatial distribution of agents; optimal in the sense that, leaving subsequently the agents forage locally and separately would be enough. In an accompanying experiment, we have implemented dispersion as an activity during which the simulated agent moves at high speed, whereas during foraging it moves at low speed (because it has to consult its sensors regularly and attentively, while foraging). However, since the critical task parameter is the spatial sample density that may take an arbitrary value, there will be eventually need for many round trips from the base to the frontiers of the explored area. Then, on the one hand the distribution would be statistically perturbed in an asynchronous manner by the agents heading to the base or returning from it, while on the other hand the distribution would have to be continuously updated as a side-effect of foraging. Moreover, such a dispersion behavior would only make sense in a world that would be small with respect to the population size, where the dispersion would ensure an efficient distribution.

To eliminate the arbitration problem between foraging and dispersion but also to allow a dynamic distribution, we introduced dispersion as an ***instrumental action*** of foraging/hunting.

Instrumental Dispersion Rule
If there is no hunting stimulus:
   If one or more agents are perceived, turn away from their local mass center (i.e., choose as travel direction the one of minimal perceived agents density).

Hunting Task with Instrumental Dispersion
If a hunting stimulus is detected,
then preferred hunting direction is the direction of the maximum perceived stimulus density,
else if one or more agents are perceived (within a certain range)

then preferred hunting direction is the direction away from their local mass center
else preferred hunting direction is the current direction
(with a small probability of random change)

We simulated the system with and without dispersion for a range of agent population sizes and the comparative results are given in Figure 10. We observe that, contrary to intuitive predictions, dispersion is suboptimal: it slightly perturbs the performance of small populations and becomes more favorable to populations of size greater than 10, giving a performance of the same order than without dispersion. This result is independent from dispersion range, i.e., from the maximal agent detection distance that triggers dispersion: we simulated the system for a dispersion range from 2 to 10 and its performance in both cases have been found comparable to this without dispersion for large populations and slightly inferior for small ones. Those two values have been chosen the first inferior and the second superior to the foraging range, i.e., the maximal sample detection distance that triggers the agent's appetitive behavior (we have fixed this range to 3, cf. the section Functional Level: Motivational System). As foreseen, the sweeping performance is worse (better) if the dispersion range is greater (smaller) than the foraging range, because in the first case, if a sample exists between two agents, they may disperse from each other and go away from it without detecting it.

This result is an indication that a reaction to a *social* stimulus *outside* the agent's motivational system does not constitute a true sociality and does not increase operationality. Instead, the above dispersion behavior
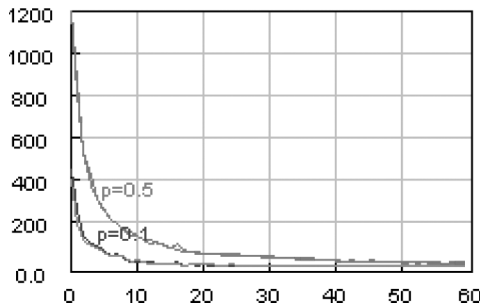

Figure 10. Task completion time with and without dispersion according to the number of explorer agents ($p_w(0) = 0.5$, dispersion range $= 3$).

is at best just part of an algorithm that makes problem solving more efficient, which should be expected from its definition as instrumental in the agent's architecture—the agent has no real motivation to flee from other agents. The behavior of the docker robot that forms robotic chains (Drogoul and Ferber 1992), the modest cooperation mechanism of Premvuti and Yuta (1990, 1993) and the architectures of cooperative or coordinated navigation (Wang 1991, Noreils 1993, Matarić et al. 1995) do not constitute a sociality either. In all those cases, the *cooperative* behavior involved is just an additional behavior that speeds up the task execution *without* intervening explicitly in the agents' motivational system. To use Matarić's (1992a) terminology, this is an "informed" agents coexistence, rather than an "intelligent" coexistence that would show as a *differential compromise between the various behaviors*, i.e., it would intervene at the motivational system level. Thus, it will make sense to talk about sociality whenever stimuli of social origin are considered by the motivational system.

Sociality Models (Cooperation Mechanisms)

Having eliminated dispersion for lack of operationality and having adopted as a starting point that sociality has to intervene at the agents' motivational system level, we sought an additional mechanism that would modify the agents' motivation taking into account the motivations of the other agents. To this end, it is necessary to introduce motivation emitters and detectors; given that motivations take continuous real values, radio signal sensors would be sufficient for a distributed robotic implementation. This sociality is *reactive* in the sense that the agents have not *discovered* it or learned it by themselves, nor have they decided that it would be better to share their needs, but it is the designer of the system that endowed them with the necessary mechanism to do so.

What motivation do the agents have to exchange? Obviously, it can neither be the foraging/hunting nor the homing motivation, but the supermotivation of p_a. We are therefore looking for new formulae of computation/adaptation of the following type:

Motivation Containing a Sociality Component

$$p_a(t) = f(p_a(t - w), p_{comp}, p_{soc})$$

where p_{comp} is computed as in the section Cognitive/Representative

Level: Recursive Satisfaction and Adaptation and p_{soc} is the perceived mean of other agents motivations.

We have implemented and compared four sociality models: (1) default *smooth* sociality, (2) relative sociality, (3) cooperative sociality, and (4) tit-for-tat sociality. The first model of reactive sociality is *smooth* sociality (its name will be explained toward the end of this section), where the agent acts as in the asocial case, starting from the social value p_{soc}.

Smooth Sociality

$$diff = |p_{comp} - p_{soc}|,$$
$$p_a(t) = p_{soc} + diff \ ^*r$$
$$\text{If } diff \leq f_p, \text{ quicker adaptation,}$$
$$\text{otherwise slower adaptation}$$

In relative sociality, the estimated sample density is given by the formula $p_{est} = a^*p_{soc} + (1 - a)^*p_{comp}$, where p_{soc} includes the agent's own value; this formula means that the estimate of the world density is considered to be somewhere between the individually perceived (p_{comp}) and the mean of agents' beliefs. The a parameter denotes the relative weight of the individual value and the estimated one. The rest of the model is identical to that of the solitary explorer agent.

Relative Sociality

$$p_{est} = a^*p_{soc} + (1 - a)^*p_{comp}, \ diff = |p_{est} - p|, \ 0 < a < 1$$
$$p = p + r^*diff$$
$$\text{If } diff \leq f_p, \text{ quicker adaptation,}$$
$$\text{otherwise slower adaptation}$$

The third model is cooperative sociality, where every agent considers that the actual value of the world density is the mean between the computed density and the mean agents density (we consider that the *truth* is the average of the two values, the personal estimate and the social standard).

Cooperative Sociality

$$p_{est} = (p_{soc} + p_{comp})/2 \ (a = 0.5), \ diff = |p_{est} - p|,$$
$$p = p_{est}$$

If diff $\leq f_p$, quicker adaptation,
otherwise slower adaptation

Finally, tit-for-tat sociality is a little more complicated than the previous two models: it is like cooperative sociality as far as world density estimation is considered, but it has furthermore a threshold used by the tit-for-tat mechanism to detect cooperation. Adaptation is thus a side-effect of cooperation or defection.

In a first nonoperational version of tit-for-tat sociality, the agent only updates its estimate during cooperation detection. On the contrary, in the second version, on cooperation the agent adopts as new estimate the new reference value (p_{est}), otherwise it adapts to that value according to the same proportional law as before, i.e., the p_a update law is *different* for cooperation and defection, but the agent adapts *continuously*.

Tit-for-tat Sociality (version 1)

$$p_{est} = (p_{soc} + p_{comp})/2 \ (a = 0.5), \ diff = |p_{est} - p|$$
If diff $\leq f_p$,
$p = p_{est}$ and quicker adaptation,
otherwise slower adaptation

Tit-for-tat Sociality (version 2)

$$p_{est} = (p_{soc} + p_{comp})/2 \ (a = 0.5), \ diff = p_{est} - p$$
If $|diff| \leq f_p$,
$p = p_{est}$ and quicker adaptation,
otherwise $p = p + f_p{}^* sign \ (diff)$
and slower adaptation

The problem with the first tit-for-tat sociality model has been that it was often leading to a stabilization of the whole population to a value of estimated density such that no agent would then modify its own estimate value and the population's activity would be maintained forever despite the exhaustion of the mineral sources; Huberman and Glance (1993) and Nowak and May (1992) presented simulation results showing precisely that spatial diversity and asynchronicity in certain distributed systems lead to global configurations either chaotic or uniform. To remedy this problem, we investigated several extensions to the cooperation criterion: noise (around 0.1), generosity in the sense of a nonzero probability of cooperation on detection of defection (around 0.1) and

self-regulation either of the selfishness a (around 0.5) either of the tolerance/threshold T (around 0.5). The self-regulation of the selfishness or the tolerance was operational as long as the meta-level (i.e., the cooperation criterion) remained *cooperative*, in the sense that detection of cooperation should induce an increase rather than a decrease of tolerance and a decrease rather than an increase of selfishness. In all cases, tolerance regulation has given a better performance than selfishness regulation (it looks that the critical parameter of the quantitative tit-for-tat model is tolerance—cf. additional results in Tzafestas 1995), but the need for a degree of noise or generosity has not been entirely eliminated. The reason appears to be that all those variants of the tit-for-tat model rely on an inactive defection (there is no action and thus no adaptation if the world is considered hostile). However, it has been argued by several authors (Ackoff and Emery 1972, Varela 1980, van Gelder and Port 1995, Tzafestas 1995) that the essence of autonomy, and for what concerns us here of operationality, is the *continuous change* of response to the world. This brought us to the actual second version of tit-for-tat sociality that defines an action for defection as well as for cooperation and that bypasses therefore that alienation problem (social stabilization, as discussed above).

The first implementation of all those models relied on localized interactions: the perceived mean of other agents' motivations was only concerning the agents inside a limited zone (a circular disc around the agent with a certain perception range, either fixed or varying across agents). This implementation yielded results hardly different from those without sociality, despite drastic quantitative improvement during the first few simulation steps. Looking closer, we observed that, since the world size (25×25) was relatively important with respect to the population size, the agents were only really socializing during the first few simulation cycles when they were dispersing away from the base and from one another. From that point on, their trajectories were diverging and their returns to the base were statistically nonsynchronized, so that practically they were meeting each other all too rarely.

The second implementation of the sociality models relied on global interactions, i.e., on a propagation of individual motivations over the whole population. We have also implemented several variants of the four sociality models according to whether social adaptation is synchronous with individual adaptation or parallel (in which case the two adaptation operations take place independently and each at its own pace) or

probabilistic (in which social adaptation takes place statistically with a certain probability). The last two variants have been found nonoperational, in that they have practically induced no performance difference with the case of asociality. The only operational possibility has remained the synchronous adaptation given before. The relative sociality model has been eliminated early in the experimentation process for lack of operationality as well: actually, relative sociality gives an estimate of the world density that has too much inertia with respect to the agents' activity and hence with respect to the true evolution of the world density (this result has been verified in the complete range of values for the parameter a from 0 to 1: the problem remained and owed its existence to the presence of the proportional factor $r^* diff$).

Figures 11a–c give the comparative results of the application of the default asocial model (i.e., that of the section Cognitive/Representative Level: Recursive Satisfaction and Adaptation) and the four social models in different world densities. The curves show clearly the performance improvement with sociality; tit-for-tat sociality is more robust and slightly less efficient than cooperative sociality in all cases (with efficiency differences being greater in lower world densities), smooth sociality comes next and relative sociality is generally much more inefficient compared to asociality.

Discussion

It is noteworthy that the transition from the solitary explorer agent's motivational system to that of the social explorer agent (any of the adaptation formulae of this paragraph, except the smooth one) is not continuous: here p_{est} depends on p_{comp} as well as on p_{soc}. The system has been simulated in the case of a single social agent in the world (in which case $p_{soc} = p$) and it has been found that it is not operational (its adaptation is too quick and irregular); consequently, an agent has to be able to *recognize* when it is alone and when it is not, in order to execute the corresponding program. We call this property *the social leap*. This observation explains the name of the smooth sociality model, which is the same for any number of agents.

The light superiority of tit-for-tat sociality with respect to cooperative sociality is due to the variation of adaptation velocities. We simulated the system with and without diversity in initial conditions ($p_a(0)$), with and without diversity in tolerance (f_p) and with and without diversity in
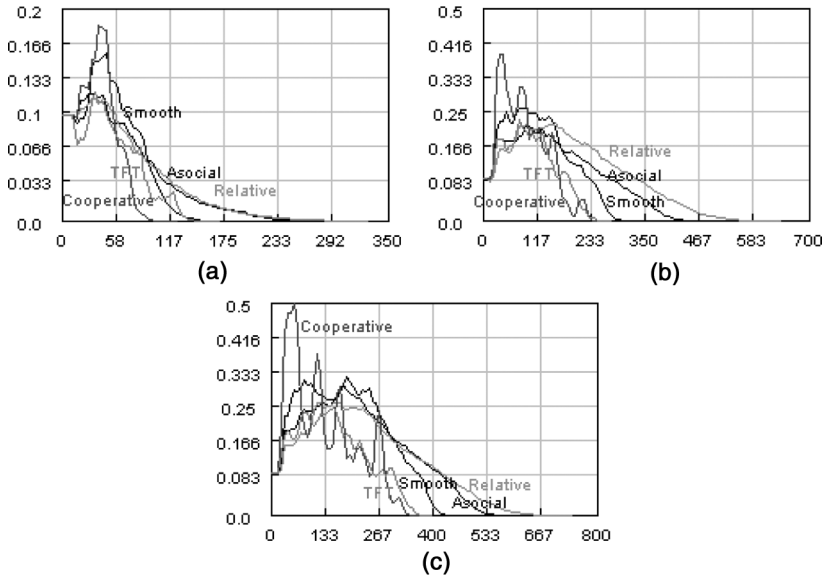
Figure 11. Asociality vs. different types of sociality in three world densities. 10 agents, (a) $p_w(0) = 0.1$, asociality $t = 336$, relative sociality ($a = 0.5$) $t = 330$, cooperative sociality $t = 127$, tit-for-tat sociality $t = 169$, smooth sociality $t = 184$. (b) $p_w(0) = 0.5$, asociality $t = 518$, relative sociality ($a = 0.5$) $t = 639$, cooperative sociality $t = 269$, tit-for-tat sociality $t = 279$, smooth sociality $t = 345$. (c) $p_w(0) = 0.9$, asociality $t = 631$, relative sociality ($a = 0.5$) $t = 743$, cooperative sociality $t = 374$, tit-for-tat sociality $t = 394$, smooth sociality $t = 475$.

adaptation parameters. Keeping initial conditions and adaptation para-meters uniform and constant in both cases, but with a tolerance diver-sity, tit-for-tat sociality has been proven slightly more operational than cooperative sociality. Why? *It appears that the tit-for-tat agents system generates diversity, since all agents do not do the same thing at the same time and this diversity is responsible for the performance improvement* (for the operational role of diversity in populations, see for example Nowak and Sigmund 1992, Lumer and Faieta 1994, Durrett and Levin 1994). However, this improvement is only marginal, since the acceleration or deceleration of adaptation does not rely on a true cooperation/defection criterion. The tit-for-tat agent is an essentially cooperative agent; it only modifies its adaptation dynamics, but it is not really defective (it has no real interest in being so, since being cooperative is not really detrimental). The need to defect arises once

we introduce factors that may make cooperation detrimental to the agent itself, for instance if we introduce *cheating agents*. We experimented with a proportion of agents that *cheat* as far as the need to communicate is concerned: more particularly we have defined agents that always emit the same constant value or a *perturbed* value (mutated with a noise probability) or a completely random value. In all cases, tit-for-tat agents have been clearly more operational than cooperative ones. However, since cheating agents do not have any particular will to cheat (it is as though their emitters were defective), there is no second *social force* (the only force present is the sweeping force) and so there is no need for compromise and sociality regulation. To show the need for sociality regulation, it would be necessary to modify the social problem so as to include a second social force, for example by defining two agent types, one that would seek to sweep the place and one that would seek to litter, or by introducing an additional social behavior such as resource sharing (as in Steels (1995) and McFarland and Spier (1997)). As the problem is formulated, there is no need for sociality regulation, because there is no free task parameter intervening in sociality.

VARIATIONS OF THE PROBLEM

Exploration of a Remote Region

Variations of the sweeping problem or extensions to conceptually similar ones may be envisaged. One such variation is the exploration or sweeping of a remote region in space, for example the exploration of the bottom right area in Figure 12. Straightforward application of the agent model to this problem is possible, but the following issues should be taken into account:

● The agents should have an additional sensing facility, allowing them to detect when they are in the target region and, if not, to compute an orientation toward it. If they lack such a facility, then the random background movement may not ensure complete coverage of the target region, because the agents may spend much time in other regions and terminate prematurely.
● If the area surrounding the remote region is empty of samples, then with appropriate parameter tuning the remote region may be completely explored and all samples exhausted.
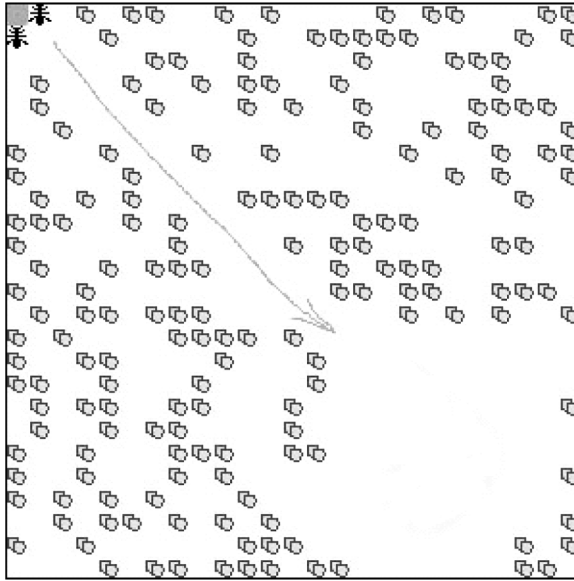
Figure 12. Exploration of a remote region (bottom right region). End state.

- If the area surrounding the remote region is not empty, then the remote region may not be totally explored and samples may not be exhausted. This is because in many cases the agents tend to get stuck in the borders of the remote region, due to detected samples outside of the region.
- To resolve this problem, we may deactivate the hunting sensor (or set it to a low range value such as 1) when the agents are outside the target area. This gives a much greater probability of proper task termination, still running the risk of missing a few remote samples, as shown in the figure. Note that when the surrounding area is not empty, an empty *corridor* from the base to the border of the region emerges, as shown in Figure 12.

Agricultural Fields

Another variation of the sweeping problem is the crop collection problem of agricultural robotics, where the swarm of agents enters the field by an edge and the shape of the field may be irregular, for example it may be rectangular as is shown in Figure 13. This environmental setting
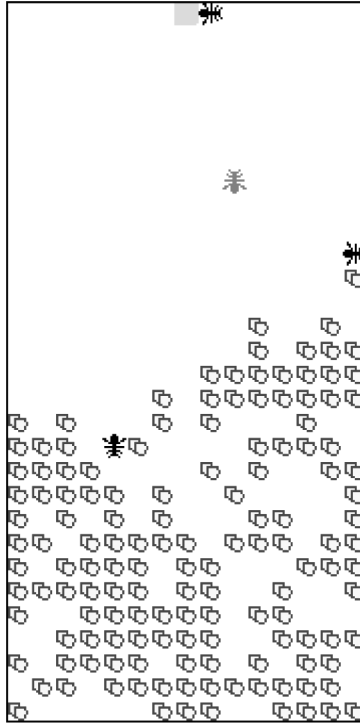
Figure 13. Exploration of a long region (15 × 30) with the entry on the smaller edge (top) by 4 agents.

induces a nonconstant task speed (i.e., a nonconstant sample picking speed), because successive trips to sample sources take longer and longer since the closer ones tend to get picked too quickly. This situation is operationally equivalent to usual circular or square worlds of very large size (for example, square worlds of size 100 × 100). In such cases, the parameter tuning of the system should be based on the worst case scenario, i.e., on the largest distance to be traveled. The resulting curve for p_a will not be bell-like as was the case in the second and third sections, but more or less irregularly decaying (see Figure 14).

DISCUSSION

Figure 15 gives the result of applying the meta-adaptation law for different initial conditions (combinations of $p_a(0)$ and $p_w(0)$). The figure
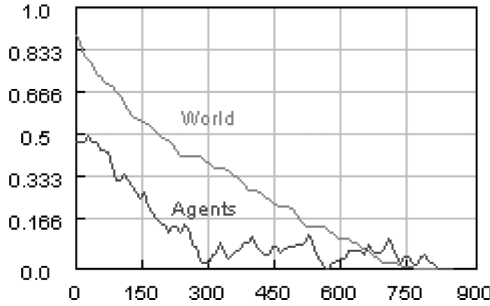
Figure 14. Coevolution agents-world for the world of Figure 13, at high density ($p_w = 0.9$).

invites the observer to think that *the initial conditions play no role in the system's operationality,* because the evolution of the curves after about 300–400 cycles is the same for all parameter settings (the three curves of each figure converge). Actually, a pair of values ($p_a(0)$, $p_w(0)$) will
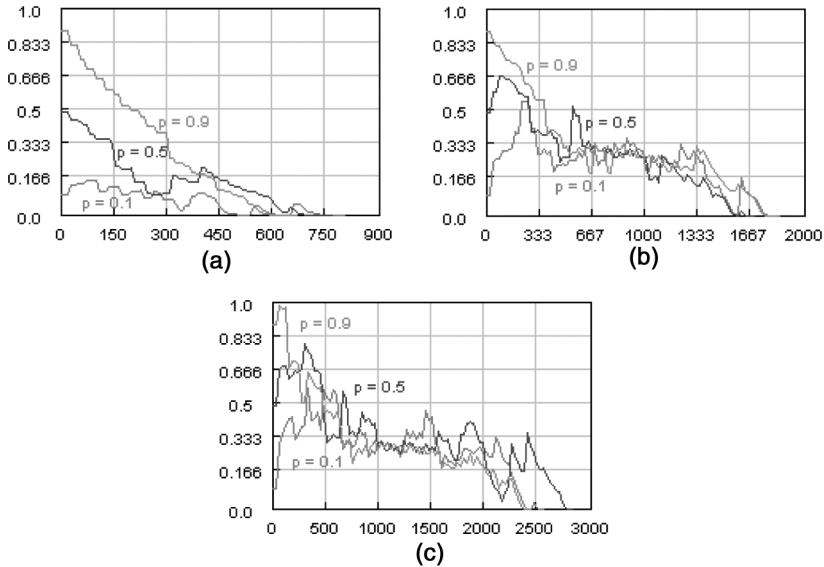

(a)

(b)

(c)

Figure 15. Performance of the agent for different initial conditions $(p_a(0), p_w(0))$. (a) $p_w(0) = 0.1$. Lower curve: $p_a(0) = 0.1$, $t_1 = 807$. Middle curve: $p_a(0) = 0.5$, $t_2 = 772$. Upper curve: $p_a(0) = 0.9$, $t_3 = 674$. (b) $p_w(0) = 0.5$. Lower curve: $p_a(0) = 0.1$, $t_1 = 1844$. Middle curve: $p_a(0) = 0.5$, $t_2 = 1698$. Upper curve: $p_a(0) = 0.9$, $t_3 = 1836$. (c) $p_w(0) = 0.9$. Lower curve: $p_a(0) = 0.1$, $t_1 = 2576$. Middle curve: $p_a(0) = 0.5$, $t_2 = 2835$. Upper curve: $p_a(0) = 0.9$, $t_3 = 2445$.

translate itself to a variable $p_a(T_0)$, where T_0 is the moment of picking the last sample ($p_w(T_0) = 0$); the study of diversity of ($p_a(0)$, $p_w(0)$) is hence equivalent to the study of diversity of $p_a(0)$ in an empty world ($p_w(0) = 0$). We have thus measured $p_a(T_0)$ for the three typical $p_w(0)$ values (0.1, 0.5, 0.9). Results are given in Table 1 for meta-adaptation. The results show that the value of $p_a(T_0)$ remain bounded below 0.2, with best match a little below 0.1.

Next, Figure 16 shows the performance of the agent for the meta-adaptation and the medium adaptation cases in an initially empty world. It may be seen that performance is almost linear in the meta case, what allows us to easily deduce the value of $p_a(0)$ from the completion time alone. More generally, we can deduce the value of $p_w(0)$ from completion time as follows:

$$t_{total} = T_0 (\text{i.e., } \textit{the time for } p_w \textit{ to go from } p_w(0) \textit{ to } 0)$$
$$+ T_1 (\text{i.e., } \textit{the time for } p_a \textit{ to go from } p_a(T_0) \textit{ to } 0)$$

The value of T_0 is a function of just $p_w(0)$, while the value of T_1 is a function of just $p_a(T_0)$. Table 1 shows that $p_a(T_0)$ is between 0 and 0.2, with maximum likelihood to be a little below 0.1. Figure 17 gives T_0 as a function of $p_w(0)$ and T_1 as a function of $p_a(T_0)$. Because both functions are linear ($T_0 = 400 + 2400p_w(0)$, $T_1 = 90 + 190p_a(T_0)$), T_1 may be given the average value of 180 and the value of $p_w(0)$ may be linearly approximated as

$$p_w(0) = c(t_{total} - t_c), \quad t_c = 580, \quad c = 1/2400$$

Table 1. $p_a(T_0)$, where T_0 = moment of last sample picking, in various ($p_w(0)$, $p_a(0)$) settings

p_a	p_w		
	0.1	0.5	0.9
0.1	0.13	0.06	0.07
0.2	0.04	0.11	0.10
0.3	0.05	0.08	0.06
0.4	0.04	0.07	0.08
0.5	0.07	0.06	0.11
0.6	0.06	0.07	0.07
0.7	0.09	0.07	0.02
0.8	0.07	0.10	0.08
0.9	0.11	0.06	0.15

Figure 16. Performance of the agent in an empty world ($p_w(0) = 0$). (a) Medium adaptation. Lower curve: $p_a(0) = 0.1$, $t_1 = 334$. Middle curve: $p_a(0) = 0.5$, $t_2 = 440$. Upper curve: $p_a(0) = 0.9$, $t_3 = 476$. (b) Meta adaptation. Lower curve: $p_a(0) = 0.1$, $t_1 = 224$. Middle curve: $p_a(0) = 0.5$, $t_2 = 466$. Upper curve: $p_a(0) = 0.9$, $t_3 = 562$.

The above formula allows us both to predict the performance of the system, i.e., for a given t_{total} compute $p_w(0)$, and to design a system that will achieve a given level of performance, i.e., to design a system whose (c, t_c) parameter pair gives a certain t_{total} for a certain initial $p_w(0)$. Figure 18 is the same as Figure 17, but for an agent having hunting range $= 2$, $r \in [0.05, 0.2]$. In this case we can verify that $T_0 = 500 + 2500 p_w(0)$, $T_1 = 100 + 400 p_a(T_0)$, thus $t_c = 800$ and $c = 1/2500$. Care has to be taken however of very sparse worlds, i.e., worlds having low initial $p_w(0)$, because $p_a(T_0)$ may vary significantly and take extreme values, even exceed 0.2 in some cases (for example in Table 1,
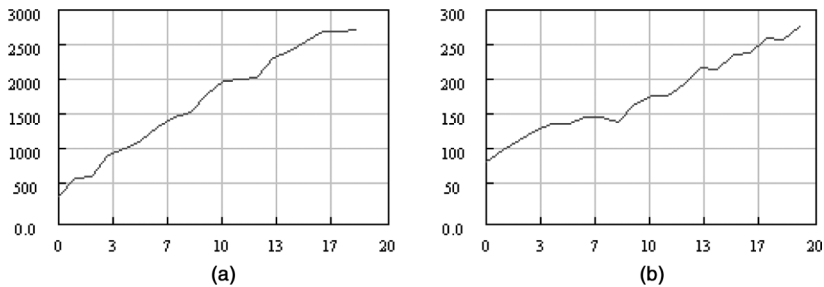

Figure 17. (a) Completion time T_0 (*i.e. time for p_w to go from $p_w(0)$ to 0*) as a function of $p_w(0)$. The x-axis represents the range of values [0, 1]. Completion time evolves almost linearly with $p_w(0)$. (b) Completion time T_1 (*i.e. time for p_a to go from $p_a(T_0)$ to 0*) of an agent in an empty world ($p_w(0) = 0$) as a function of $p_a(0)$. The x-axis represents the range of values [0, 0.2]. Completion time evolves almost linearly with $p_a(0)$.
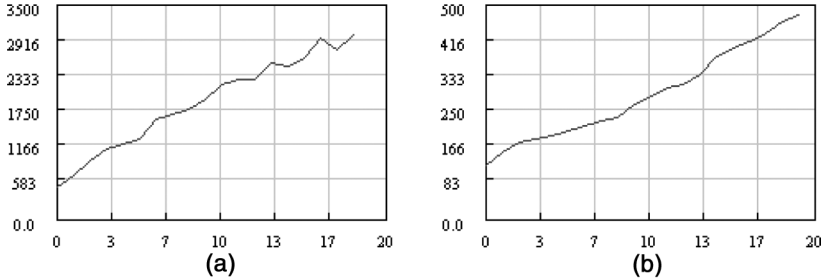
Figure 18. Agent having hunting range $= 2$, and $r \in [0.05, 0.2]$. (a) Completion time T_0 *(i.e., time for p_w to go from $p_w(0)$ to 0)* as a function of $p_w(0)$. The x-axis represents the range of values $[0, 1]$. Completion time evolves almost linearly with $p_w(0)$. (b) Completion time T_1 *(i.e., time for p_a to go from $p_a(T_0)$ to 0)* of an agent in an empty world $(p_w(0) = 0)$ as a function of $p_a(0)$. The x-axis represents the range of values $[0, 0.2]$. Completion time evolves almost linearly with $p_a(0)$.

$p_a(T_0) = 0.13$ for $p_a(0) = p_w(0) = 0.1$). This may be explained because for low initial $p_w(0)$, T_0 obtains a value close to its minimum, which may belong to the range of values of T_1. An alternative explanation is that the system does not have sufficient time to become initial value independent. As has been shown in the beginning of this section for Figure 15 the necessary time is around 300–400 cycles for the default case of Figure 17, and should be even greater for the agent of Figure 18 that adapts slower on average.

As a conclusion of the above discussion, the analysis of any such explorer system may be done with reference to the boundary conditions $p_w(0) = 0$ and $p_a(0) = 0$. The agent's individuality that is manifest in its responses to the various perturbations is precisely its response to the boundary conditions. In other words, and thanks to behavioral linearity, to isolate the difference between two agents having different behavioral and adaptation parameters, it is sufficient to study them in this degenerate case instead of studying them exhaustively in all possible perturbation cases.

CONCLUSION

In this paper we studied the control system of an explorer-sweeper agent that has to visit and exhaust all sources of interest within a delimited area. It has been shown that an agent with two tasks that express coupled motivations, a recursive motivation and representation of the critical free task

parameter constitutes a simple solution to the sweeping problem. The state of the world in which the agent is situated is a source of persistent perturbation for the agent, who has to modify its dynamics of interaction with the world in order to resolve his (our) problem better. The modification of the interaction dynamics concerns the adaptation rates of the agent and has to be endogenous, homeostatic, and bang-bang in order to ensure the best operational coupling between agent and world.

However, and since the agent has fixed motivations, representations and ways of coupling with the environment, we might say that it is not autonomous in the sense of being self-determined, but it acts as our controller for the sweeping problem. The individuality of that controller is seen in its response to the boundary conditions $p_w(t) = 0$ and $p_a(t) = 0$, that is the study and analysis of the controller may take place outside a particular environment and a particular interaction context. Thus, the agent is autonomous in the sense that it does not depend on the interaction, but it acts thanks to internal designed motivations, according to internal laws and independently of its environment that only perturbs him temporarily. Table 2 summarizes our results for the case of the solitary agent.

The sweeping problem has also been studied in the case of multiple agents instead of a single agent. First, we studied the performance of the multiagent system as a function of the population size and we showed that the performance comes to a saturation level. Next we studied the effect of the introduction of an additional dispersion possibility as an

Table 2. Recapitulation of single agent case

The problem	Control system of an agent needing to *know* the critical parameter of the problem it solves and whose value denotes the state of progress of the solving process
The application	Solitary explorer agent
The solution	*Motivational system* (hunting and homing motivations)
	Representational variable
	Estimation of the variable (adaptation)
	Self-regulation of the dynamics of interaction with the world (adaptation parameters)
Issues	● Endogenous versus exogenous adaptation
	● Gradual versus bang-bang dynamics
	● Operational coupling (meta-adaptation)
	● Insensibility to initial conditions
	● Predictability of performance

Table 3. Recapitulation of multiple agents case

The problem	Control system of an agent having to carry out its task in common with other agents
The application	Social explorer agent
The solution	*Solution for solitary explorer agent + Sociality* (during update of the representational variable take into account the other agents' estimations as well)
Issues	• Saturation of the populations' performance
	• Dispersion ⇒ Slight decrease and instability of performance
	• Smooth sociality vs. cooperative sociality vs. relative sociality vs. tit-for-tat sociality
	• A case of nonoperationality because of stabilization of the dynamics of interaction
	• The social leap
	• The role of diversity
	• Presence of cheater agents

instrumental action for foraging/hunting. This time, we observed a slight decrease in efficiency, due to the instability of the induced dynamic spatial structures. The implementation and the comparison of several models of reactive sociality led us to the conclusion that we need a tit-for-tat sociality of the type *social variable sharing*. We also presented other assorted observations, the social leap phenomenon, the role of diversity and the behavior of the system in presence of cheating agents. Table 3 summarizes our results for the case of the social explorer.

In the future, we purport to extend this basic behavioral system to include an energy managing task, that will drive the agent(s) to recharge at the same or at a different base, so as to study potential new intricacies of the motivational and the adaptation system. Yet another application envisaged is the sweeping for more than one different types of object, where inspiration will be drawn from usual ant clustering (Martinoli et al. 1999) and division of labor (Théraulaz et al. 1998) approaches and results. Finally, we plan to model and study populations of two behaviorally mutually exclusive agent types.

ACKNOWLEDGMENTS

REFERENCES

Ackoff, R. L. and F. E. Emery. 1972. *On purposeful systems*. Chicago: Aldine Atherton.

Angle, C. M. and R. A. Brooks. 1990. Small planetary rovers, *Proceedings IEEE International Workshop on Intelligent Robots and Systems (IROS) '90*, Japan, 6 pages.

Arkin, R. C., T. Balch, and E. Nitz. 1993. Communication of behavioral state in multi-agent retrieval tasks, *IEEE International Conference on Robotics and Automation*, Atlanta, GA, May, pp. 588–594.

Beckers, R., O. E. Holland, and J.-L. Deneubourg. 1994. From local actions to global tasks: Stigmergy and collective robotics, *Artificial Life IV, Proceedings of the 4th Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (R. Brooks and P. Maes, Eds.), MIT Press, Cambridge, MA: pp. 181–189.

Beer, R. D. and H. J. Chiel. 1990. Neural implementation of motivated behaviour: Feeding in an artificial insect. In *Advances in neural information processing systems 2*, edited by D. Touretzky. Morgan Kaufmann, pp. 44–51.

Brooks, R. A. and A. M. Flynn. 1989. A robot being. In *Robots and biological systems: Towards a new bionics*? edited by P. Dario, G. Sandini, and P. Aebischer, pp. 679–701.

Deneubourg, J.-L. , S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chrétien. 1990. The dynamics of collective sorting—Robot-like ants and ant-like robots. In *Animals to animats, Proceedings of the 1st International Conference on Simulation of Adaptive Behavior* (J.-A. Meyer and S. W. Wilson, Eds.), Bradford/MIT Press, and Cambridge, MA: pp. 356–363.

Drogoul, A. and J. Ferber. 1992. From Tom Thumb to the dockers: Some experiments with foraging robots. In *Animals to animats 2, Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior* (J.-A. Meyer, H. L. Roitblat, and S. W. Wilson, Eds.), Hawaii: Bradford/MIT Press, and Cambridge, MA: pp. 451–459.

Durrett, R. and S. A. Levin. 1994. The importance of being discrete and spatial. *Theoretical Population Biology*, 46(3):363–394.

Huberman, B. A. and N. S. Glance. 1993. Evolutionary games and computer simulations. *Proceedings National Academy of Sciences (USA)*, Vol. 90, pp. 7716–7718.

Kube, C. R. and E. Bonabeau. 2000. Cooperative transport by ants and robots. *Robotics and Autonomous Systems*, 30(1–2), pp. 85–101.

Lumer, E. D. and B. Faieta. 1994. Diversity and adaptation in populations of clustering ants. In *Animals to animats 3, Proceedings of the 3rd International Conference on Simulation of Adaptive Behavior* (D. Cliff, P. Husbands, J.-A. Meyer, and S. W. Wilson, Eds.), Bradford/MIT Press, Cambridge, MA: pp. 501–508.

Maes, P. 1990. A bottom-up mechanism for action selection in an artificial creature. In *Animals to animats, Proceedings of the 1st International Conference on Simulation of Adaptive Behavior* (J.-A. Meyer and S. W. Wilson, Eds.), Paris, Bradford/MIT Press, pp. 238–246.

Maes, P. 1991. Learning behavior networks from experience. *Toward a practice of autonomous systems, Proceedings of the 1st European Conference on Artificial Life* (F. J. Varela and P. Bourgine, Eds.), Paris, December, MIT Press/ Bradford Books, pp. 48–57.

Martinoli, A., A. J. Ijspeert, and F. Mondada. 1999. Understanding collective aggregation mechanisms: From probabilistic modelling to experiments with real robots. *Robotics and Autonomous Systems*, 29(1):51–63.

Matarić, M. J. 1992a. Minimizing complexity in controlling a mobile robot population, *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, Nice, France, May, pp. 830–835.

Matarić, M. J. 1992b. Designing emergent behaviors: From local interactions to collective intelligence. In *Animals to Animats 2, Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior* (J.-A. Meyer, H. L. Roitblat, and S. W. Wilson, Eds.), Hawaii, Bradford/MIT Press, pp. 432–441.

Matarić, M. J. 1994. *Interaction and intelligent behavior*, Ph.D. Thesis, MIT Artificial Intelligence. Laboratory Technical Report AITR-1495.

Matarić, M. J., M. Nilsson, and K. Simsarian. 1995. Cooperative multi-robot box-pushing, *Proceedings IROS-95*, Pittsburgh, PA: pp. 556–561.

McFarland, D. and T. Bösser. 1993. *Intelligent behavior in animals and robots*. Cambridge, MA: MIT Press.

McFarland, D. and E. Spier. 1997. Basic cycles, utility and opportunism in self-sufficient robots. *Robotics and Autonomous Systems*, 20(2–4), pp. 179–190.

Noreils, F. R. 1993. Toward a robot architecture integrating cooperation between mobile robots: Application to the indoor environment. *International Journal of Robotics Research*, 12(1):79–98.

Nowak, M. A. and R. M. May. 1992. Evolutionary games and spatial chaos. *Nature*, 359:826–829.

Nowak, M. and K. Sigmund. 1992. Tit-for-tat in heterogeneous populations. *Nature*, 355:250–253.

Premvuti, S. and S. Yuta. 1990. Consideration on the cooperation of multiple autonomous mobile robots, *Proceedings 1990 IEEE International Workshop on Intelligent Robots and Systems (IROS '90)*, Japan, pp. 59–63.

Premvuti, S. and S. Yuta 1993. An experiment in realizing autonomous and cooperative behaviours between multiple mobile robots. *IEEE International Workshop on Emerging Technologies and Factory Automation (Technology for the Intelligent Factory)*, R. Zurawski and T. S. Dillon (Eds.), Melbourne, Australia, pp. 301–304.

Steels, L. 1990. Cooperation between Distributed Agents through Self-Organisation, *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back* (P. Maes, Ed.), MIT Bradford Press, reprinted from Special Issue of the *Journal on Robotics and Autonomous Systems*, 6(1&2), pp. 71–80.

Steels, L. 1995. Discovering the competitors, *Adaptive Behavior*, 4(2):173–199.

Théraulaz, G., E. Bonabeau, and J.-L. Deneubourg. 1998. Response threshold reinforcement and division of labour in insect societies, *Proceedings of the Royal Society of London B*, 256:327–332.

Tyrrell, T. 1993a. *Computational mechanisms for action selection*, Ph.D. Thesis, Centre for Cognitive Science, University of Edinburgh.

Tyrrell, T. 1993b. The use of hierarchies for action selection. *Adaptive Behavior*, 1(4):387–420.

Tyrrell, T. 1994. An evaluation of Maes's bottom-up mechanism for behavior selection. *Adaptive Behavior*, 2(4):307–348.

Tzafestas, E. S. 1995. Vers une systémique des agents autonomes: Des cellules, des motivations et des perturbations. *Thèse de Doctorat de l'Université Pierre et Marie Curie*, Paris.

Tzafestas, E. S. 2001. Aging agents, *Artificial Life and Robotics*, 5:46–57.

van Gelder, T., R. Port. 1995. It's about time: An overview of the dynamical approach to cognition. In *Mind as motion: Explorations in the dynamics of cognition*, edited by R. Port & T. van Gelder, Eds. Bradford/MIT Press.

Varela, F. J. 1980. *Principles of biological autonomy*, New York: Elsevier/North Holland.

Wang, P. K. C. 1991. Navigation strategies for multiple autonomous mobile robots moving in formation. *Journal of Robotic Systems*, 8(2):177–195.