Tom Thumb Robots Revisited : Self-Regulation as the Basis of Behavior

Elpida S. Tzafestas

Intelligent Robotics and Automation Laboratory
Electrical and Computer Engineering Department
National Technical University of Athens
Zographou Campus
Athens 15773, GREECE
brensham@softlab.ece.ntua.gr

Abstract

We analyze the problem of the Tom Thumb robots, i.e., of robots that forage in a closed world and communicate the position of sources using a crumb laying technique. We demonstrate that past solutions to the problem suffer from physical instability, due to crumb exhaustion for individual robots, and we propose as solution a self-regulating mechanism. On a second level, we demonstrate that the introduction of an additional self-regulation loop, parallel to the first, improves the performance of the system. A number of theoretical conclusions are drawn, the most prominent being that the actual collaborative behavior of the system is the by-product of a self-regulation process within each of the agents and that the second regulation loop concerns the parameters that define the temporal dynamics of behavior.

1 Introduction

One classical problem on the intersection of artificial life and behavior-based robotics is the robot foraging problem, where one or more robots forage locally for some source of interest, such as food or minerals. In the usual version of the problem (Steels 1990, Mataric 1992, Drogoul and Ferber 1992) there are a few large sources distributed in the world, while in (Tzafestas 1995) we have tackled the case of more or less uniform source distribution. The solution to the usual case consists in allowing a robot to lay down trails or "crumbs" while carrying a source sample to a home base, that another robot or itself may follow to arrive to the source quickly. A variant of the problem considers that trails laid down by the robots evaporate slowly, in the same way as pheromone quantities laid down by real ants in the physical world (Deneubourg et al. 1990, Nakamura and Kurumatani 1996).

We reexamine the usual version of the problem from a different point of view, in an attempt to identify or specify the conditions of validity of the solution found in the literature. The most complete solution to date has been given in (Drogoul & Ferber 1992), where a number of increasingly complex and increasingly satisfactory solutions have been analyzed. The Tom Thumb robot is able to successfully build, reinforce and correctly use trails from the home base to the source, while the Docker robot (Drogoul & Ferber 1992) uses an additional mechanism of sample "theft" from neighbors, which allows robots to build chains resembling harbor Dockers. The motivation for our work has been our feeling that the Tom Thumb robot as defined is not stable because it assumes unbounded numbers of "crumbs", which is not physically possible, and which would show in a real robotic implementation.

2 Why Tom Thumb robots fail

The Tom Thumb robot's behavioral diagram as described in (Drogoul & Ferber 1992) is depicted in Figure 1.

```
If (carrying samples)
    If (back home) lay down samples
    Else {go home, lay down 2 crumbs}
Else
 If (found samples) pick up samples
 Else
    If (crumb or stimulus sensed)
       {follow stimulus, pick up 1 crumb}
    Else move randomly
```

Figure 1. The behavioral diagram of the Tom Thumb robot (cf. Drogoul & Ferber 1992, p. 455). In the Docker robot, the condition (crumb or stimulus sensed) is replaced by (crumb or stimulus or loaded robot sensed).

The Tom Thumb robot lays down two crumbs while homing, and picks up one crumb while following crumbs or stimuli. Unless otherwise stated, all simulations reported below use a 30x30 grid world with the home base in the center emitting an orientation signal, a large source at one of the corners and a population of 10 robots starting with 50 crumbs each. Robots may sense a sample or crumb from a distance of up to three grid cells.

We have simulated first the behavior of the system as is, by measuring the quantities of crumbs deposited in the world or owned by individual agents. The results are given in figures 2 and 3. As was expected, the quantities of crumbs owned by robots generally fall below zero, while the quantity of crumbs deposited in the world may rise without limit. The exact values of these quantities depend on the problem parameters (distance from source to home base, number of robots and source size) that define the expected number of robot trips source-base necessary to complete the task.

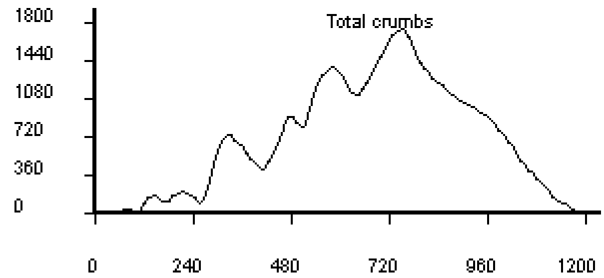

Figure 2. Quantity of crumbs in the world in a typical run (the maximum is around 1400 crumbs, which is much more than the total number of crumbs owned by all robots). The job is over when the source is exhausted and all crumbs are collected, i.e., when the path to the source has vanished.

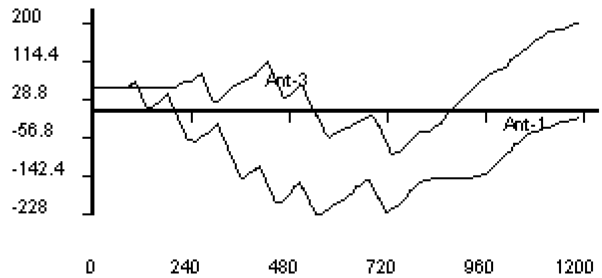

Figure 3. Quantities of crumbs owned by two robots in the above run. Both fall below zero.

An apparent question arising at this point is, "what if we just constrain robot behavior so as not to lay down crumbs when it does not have any ? aren't crumbs deposited so far enough ?" We have been able to see in several experiments that, first, depending on the problem parameters, the total quantity of crumbs might not be sufficient, in which case the path to the source will be disconnected, and, second, when it is sufficient — for instance if we start the above experiment with 1000 crumbs per agent — the total number of crumbs deposited in the world may rise tremendously. This last condition generates an important problem : the robots will continue being attracted for a long time to an empty source, that is, the surplus crumbs will be misleading. This observation brings us to the actual formulation of the above trailing problem :

We are seeking a laydown-pickup mechanism such that a trail to a source is built quickly and reinforced while the source exists and vanishes shortly after the source is exhausted.

3 The solution : Self-regulation

The problem of agent crumb exhaustion lends itself to a simple solution. Every time a robot needs to lay down or pick up crumbs, it should do it in a way so as to preserve its own quantity of crumbs within some desired bounds $crumbs_{min}$ and $crumbs_{max}$, by using the following laws :

For laydown (1a)

$$crumbs(t+1) = crumbs(t) + r_l * (crumbs_{min} - crumbs(t))$$

For pickup (1b)

$$crumbs(t+1) = crumbs(t) + r_p * (crumbs_{max} - crumbs(t))$$

This simple regulation mechanism ensures that no agent will ever run out of crumbs completely. However, the absolute (real-valued) quantity of crumbs deposited or collected at each cycle will depend on the state of the agent : an agent with many crumbs will lay down more and pick up less than an agent with just a few crumbs remaining. This arrangement allows for trails to be built rapidly (because agents in the beginning have a statistically medium number of crumbs, so they tend to lay down large quantities of crumbs) and to vanish quickly (because agents toward the end of the task have statistically only a few crumbs, so they tend to pick up large quantities of crumbs). In what follows it will be assumed that $crumbs_{min}$=10 and $crumbs_{max}$ = 100, for all agents.

4 Meta-regulation : Temporal dynamics

While we can certainly fix r_l and r_p to two values and get the system running, it is an important concern to identify proper values for these parameters, i.e., values that will ensure a "statistically optimal" performance, according to the problem formulation given at the end of section 2. Intuitively, and all other things being equal, we expect to have different "optimal" values of r_l and r_p, for different environmental conditions. In figure 4 we give the comparative results of a typical simulation run with r_l = 0.12 and r_p=0.06 for three cases of a small, a medium and a large source size (20, 50 and 80 samples, respectively).

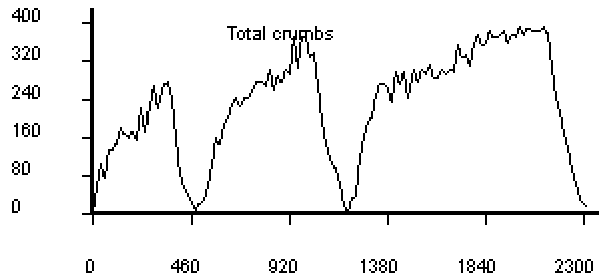

Figure 4. Comparative performance for a typical simulation run with r_l=0.12, r_p=0.06 in three environments where the source size is 20, 50 and 80, respectively. The duration of the task is 488, 696 and 1105 cycles, respectively.

We have conducted experiments with various parameter settings in various environmental conditions and we have obtained results that differ both quantitatively and qualitatively. However, all of these parameter settings share the essential characteristic of uniform laydown or pickup rates. A large laydown rate will be beneficial in the start and middle of the task, when the agents would like to build and reinforce a trail quickly, while a large pickup rate would be beneficial toward the end of the task, when the agents would like to destroy the trail to the exhausted source as quickly as possible. While a given parameter setting would be more desirable than another one in a particular context, our goal as designers should be to ensure the better behavior *globally*, i.e., to ensure that the system will "discover" or identify the proper parameter setting in each situation.

Consequently, what we really want is *not* a particular parameter setting, but a mechanism that will allow a robot to lay down more and pick up less crumbs at the beginning of the task (so as to build and reinforce the path) and vice versa toward the end (so as to destroy it quickly). To this end, a measure of the state of the task must be available. The only such measure that a robot may have is the number of the crumbs in the world. However, since this quantity cannot be directly perceivable, we have used an estimate of it, simply the number of crumbs at the current position of the robot. This estimate is used as follows :

For laydown

If crumbs(t) >= world_crumbs_estimate (2)

$$r_l(t+1) = r_l(t) + r_{rl} * (r_{lmax} - r_l(t))$$

else

$$r_l(t+1) = r_l(t) + r_{rl} * (r_{lmin} - r_l(t))$$

For pickup

If crumbs(t) >= world_crumbs_estimate (3)

$$r_p(t+1) = r_p(t) + r_{rp} * (r_{pmin} - r_p(t))$$

else

$$r_p(t+1) = r_p(t) + r_{rp} * (r_{pmax} - r_p(t))$$

As is obvious from the formulae, the rate of crumb laying increases when the robot owns more crumbs than may be found in its current position and decreases otherwise. Inversely, the rate of crumb picking increases when the robot owns less crumbs than may be found in its current position and decreases otherwise.

Figure 5 gives the result of the application of the above model in the three environmental settings used in figure 4. Surprisingly enough, the self-regulation of the laydown and pickup rates not just does change the shape of the curves, i.e., the qualitative behavior of the agents (the quantity of crumbs in the world rises quickly to a fairly high value, stays close to it during the task, and falls back quickly to zero when the source is exhausted, while showing far less fluctuations than in the previous case), but it improves results quantitatively as well : in all runs, including the one depicted, the duration of the task has been shorter than with the previous model.

Figure 6 gives the curves of the r_l and r_p parameters of one of the agents in the above run. It is clearly seen that r_l is high at the beginning and during foraging, while r_p is high toward the end of the task.

This improvement is more pronounced in harder environments where the regulation needs are more urgent, for example in the case of longer distances from home base to source or in the case of more agents. Also, figure 7 gives comparative results without and with meta-regulation for the case of Docker robots. Note that the performance is inferior to the one of Tom Thumb robots (middle curve of figure 5). This result is most probably statistically insignificant, but the actual comparative performances for Tom Thumb robots and Dockers in the case of meta-regulated behavior remain to be investigated. Note also that in this last setting there are more fluctuations in the shape of the path than in the previous ones, because since Dockers "steal" samples from one another the crumbs path are generally neither continuous nor persistent.

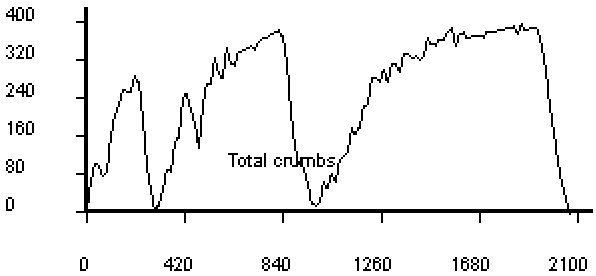

Figure 5. The same experiment as in figure 4 but with meta-regulation of r_l and r_p, between 0.06 and 0.12 for each one of them ($r_{rl} = r_{rp} = 0.1$). The task duration is 302, 676 and 1063 cycles, respectively. The maximum number of crumbs in the world is approximately the same in all three cases, because they are laid down quickly enough, and on average higher than in the previous cases.

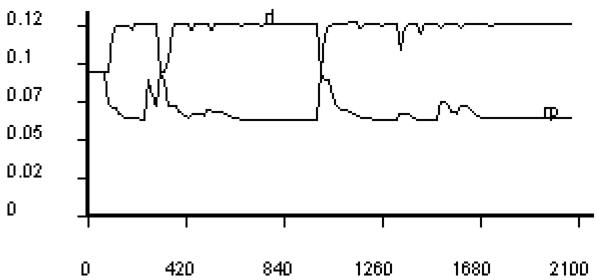

Figure 6. Curves of r_l and r_p, for an agent in the run of the previous figure. Local peaks of r_p correspond to situations where the agent has had to return to the home base while the trail was temporarily disrupted. However, the agent has been able to return to the correct behavior quickly. Similar observations may be drawn for r_l. Notice that the value of r_l only changes during the beginning of the task, when agents collect samples and need to lay down crumbs.

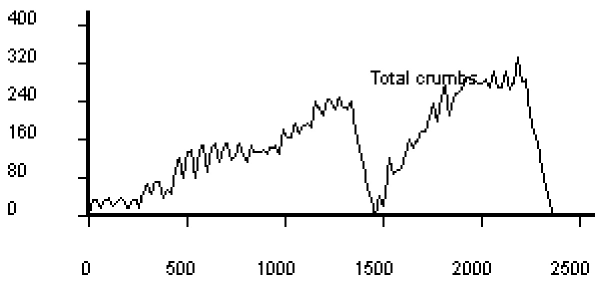

Figure 7. Comparative results without and with meta-regulation for the case of the agents having the Docker behavior as described in (Drogoul and Ferber 1992). The task duration is 1440 and 1021 cycles, respectively.

5 Theoretical discussion

We have shown above that the agent's behavior is based on a critical variable (the individual crumb quantity) that drives its motivation to participate in the crumb laying and picking process. This variable is coupled with the actual quantity of crumbs in the environment through the agent's behavior. By regulating its own variable, an agent tries to bring the corresponding world variable to 0. In (Tzafestas 1995) we have called this property of the agent-world system "*operational coupling*", since it defines a coupling between agent and world such that the agent's behavior is qualitatively operational, that is, it responds to the environmental perturbations in a uniform way. Furthermore, this variable has *cognitive value*, since it represents the agent's idea about the state of the environment (a low value of the agent variable most probably means a world where a source exists). Seen this way, the agent may be thought of as trying to approach or approximate the world variable, i.e., as trying to adapt to its environment.

The operationality of the behavior is ensured through an additional self-regulation mechanism acting on the adaptation rates. This is an important observation, since it is compatible with the dynamical approach to cognition (van Gelder and Port 1995), stating that the most important factor in cognitive mechanisms is the nature of dynamics involved. Mechanisms like the ones developed here may be also regarded as a first step toward the realization of autopoietic systems :

> "… an autopoietic system is a homeostat … the critical variable is *the system's own organization.* It does not matter, it seems, whether every measurable property of that organizational structure changes utterly in the system's process of continuing adaptation. *It* survives." (Maturana and Varela 1980, p. 66, authors' emphasis)

Of course, we have explored many unsuccessful regulation variants as well, the most important being the inverse regulation scheme, where in formulae (2) and (3) the inequalities are inversed. A comparison of the two mechanisms showed that the inverse mechanism is unsuccessful because agents then take the environment's state into account negatively, so that they appear non cooperative to other agents. For instance, an agent possessing many more crumbs than there are in its environment will try to give away as little or possible or pick up as many as possible, so as to maintain this difference, hence hiding information from other agents. Of course, this kind of behavior will have a negative impact on itself as well, because if other agents do not find a path to a source, he won't either. This is another demonstration of the well-known principle that cooperative behavior is first of all selfish (Axelrod 1984).

The final observation concerns the point of view taken to analyze this problem. While it has been traditionally tackled as an engineering problem, where the goal has been to solve a primitive problem of communication between agents, in this work we are proposing an inverse point of view, where the agent may be thought of as trying to regulate within bounds some internal variables (the regulated variables appear to be critical for an agent's survival or operationality, so that Ashby (1960) calles them *essential variables*). The buildup and reinforcement of the communication means, i.e., of the trail, is a by-product of agent self-regulation when a perturbation occurs, i.e., when sample sources exist. The driving force of the agent's behavior is thus the state of its essential variables, whereas the picking and laying components constitute the metabolic part of the overall mechanism.

It is noteworthy that exactly the same qualitative conclusions have been drawn in the case of agents exploring an environment with more or less uniform distribution of sources (Tzafestas 1995), though with a different cognitive variable and a different type of first-level adaptation.

6 Conclusions and perspectives

We have investigated the classical robot exploration problem in the case of a few large localized sources and we have shown that the fundamental Tom Thumb solution is not complete from a physical and stability point of view, since individual agents run out of crumbs or the world gets overwhelmed with unnecessarily large quantities of them. What is necessary is a regulation mechanism that ensures that no agent will fall out of bounds as far as its own quantity of crumbs is concerned. The regulation model yields a better performance than the original Tom Thumb solution. On top of that, a second regulation loop is introduced that acts on the rates of the first one. The meta-regulation mechanism improves the performance of the agents and this improvement is more pronounced in harder problems where the regulation needs are more urgent, such as longer distances home-source, or larger numbers of agents, or Docker behavior. Theoretically, the overall model relies on the definition of a *cognitive variable* for each agent, that is coupled with an environmental variable and is adapted by the agent throughout the job. The adaptation rates that define the dynamics of the system are themselves regulated within bounds and this constitutes the meta-regulation loop. Overall, the agents may be regarded as self-regulating some internal "essential" variables, with the by-product being the communication with other agents through trails and the completion of the task.

The linear regulation model is by no means new. It is a fundamental model in early cybernetics research and it is also widely used in reinforcement learning work. Note, however, that our problem is *not* a learning one, in the usual sense of the term. To our opinion, this is an indication that the basic mechanisms underneath learning (be it linear regulation or others) preexist in an agent for some other reason, namely to solve some more primitive adaptation problems before true learning becomes necessary.

In the past, our approach has been already validated for the exploration problem in a uniform source distribution and the same principles have been found to apply. The next step is to formulate and solve in the same way a few other classical artificial life problems, such as robot cooperation in a closed ecosystem (Steels 1994) and action selection (Tyrrell 1993,1994). We hope that the comparative study of the results and conclusions for each of these problems will reveal a few secret principles for engineering or understanding regulation mechanisms.

Acknowledgments. Most of the ideas presented in this paper emerged while I was with the LAFORIA-IBP at Université Pierre et Marie Curie, Paris, France. I wish to thank Professor Jacques Ferber, as well as Claude Delaye, Steffen Lalande and Stéphane Bura, for inspiring discussions.

References

Ashby, W.R. 1960. *Design for a brain - The origin of adaptive behaviour.* 2nd revised edition, London: Chapman & Hall.

Axelrod, R. 1984. *The evolution of cooperation.* Basic Books.

Deneubourg, J.-L., Aron, S. Goss, S. and Pasteels, J.M. 1990. The self-organizing exploratory pattern of the Argentine Ant. *Journal of Insect Behavior* 3:159-168.

Drogoul, A. and Ferber, J. 1992. From Tom Thumb to the Dockers : Some experiments with foraging robots. In Proceedings Simulation of Adaptive Behavior 1992, 451-459.

Mataric, M. 1992. Designing emergent behaviors : From local interactions to collective intelligence. In Proceedings Simulation of Adaptive Behavior 1992, 432-441.

Maturana, H.R., and Varela, F. 1980. *Autopoiesis and cognition — The realization of the living*, Dordrecht/ Boston, D. Reidel Publishing.

Nakamura, M. and Kurumatani, K. 1996. Formation mechanism of pheromone pattern and control of foraging behavior in an ant colony model. In Proceedings Artificial Life V, 67-74.

Steels, L. 1990. Towards a theory of emergent functionality. In Proceedings Simulation of Adaptive Behavior 1990, 451-461.

Steels, L. 1994. A case study in the behavior-oriented design of autonomous agents. In Proceedings Simulation of Adaptive Behavior 1994, 445-452.

Tyrrell, T. 1993. The use of hierarchies for action selection. *Adaptive Behavior* 1:387-420.

Tyrrell, T. 1994. An evaluation of Maes' bootom-up mechanism for behavior selection. *Adaptive Behavior* 2:307-348.

Tzafestas, E. 1995. Vers une systémique des agents autonomes : Des cellules, des motivations et des perturbations. Ph.D. diss., LAFORIA-IBP, Univ. Pierre et Marie Curie, Paris.

van Gelder, T., and Port, R. 1995. It's about time : An overview of the dynamical approach to cognition, in *Mind as Motion* : *Explorations in the dynamics of cognition*, by T. van Gelder and R. Port, Cambridge, Mass.: MIT Press.