# On the Development of a Swarm Engineering Methodology

S. Kazadi
Jisan Research Institute
28 N Oak Ave
Pasadena, CA 91107

## Abstract

This paper explores swarm engineering by revisiting popular concepts from swarm intelligence and making them more rigorous by providing mathematical definitions. The definitions form the basis for an examination of an engineering methodology which starts by examining the desired state of a global property of the system and then generates a requirement for a local behavior that will generate the global property. This methodology allows a local behavior to be tested theoretically before it is tested empirically.

**keywords**: swarm engineering

## 1 Introduction

In recent years many researchers have studied swarms in the hope of utilizing their properties to solve varied engineering problems. This field has grown into a largely experimental field centered around the application of basic principles to problems in various fields including construction, optimization, computation, and distributed sensing [Bonabeau 1999]. The field, known as swarm intelligence, was initially motivated by studies which indicated that intelligent behavior could result from the group dynamic. Swarm intelligence has recently developed into a more heavily explored field, encompassing many more potential uses. As this has happened, several themes have appeared which seem to be central to swarm-based design. These themes, which will not be exhaustively reproduced here, are forming the basis for an entirely new and less ad-hoc science.

In addition to the increasing number of themes, it has become clear recently that more rigorous methodologies are needed in order to provably design complex swarms capable of carrying out specific tasks. It would seem to be necessary to explore methods of *engineering* swarms with provable properties. What started out as an intellectual curiosity is rapidly becoming an engineering field, and it is at this point that methods of this *swarm engineering* be carefully designed. While many of the earlier studies that have looked at swarm behaviors have been able to describe the system behavior, few were able to explain the general control mechanism at work or to give clues about how the ideas generated could be generalized to other swarms. As a result, most of the work done in swarms has centered around the application of a small number of swarms that are already rather well-understood to various problems which can be formulated in similar fashions.

However, in order to generate a truly lasting science, analytic methods must be developed which allow swarms to be understood predictively. While several studies [Goldberg 2000, Spears 2004, Jones 2004] have explored the behavior of specific swarms, very few studies have been predictive in the sense that they give clues about how one might design swarms *a priori*. A recent example may be found in the work of [Lerman 2001] in which very accurate predictions about the behavior of the swarm were made, but it was not immediately clear how this might be applied to other swarms of dissimilar agents. [Kazadi 2004, Spears 2004, Jones 2004] describes some of the dynamics of puck clustering, and this is certainly predictive for clustering systems, but this still lacks the general methodology required to design arbitrary swarms with large numbers of different purposes and properties.

In this paper, we explore the development of a firm set of definitions and procedures that lead one to the development of a swarm of agents specific to a given global task. Our hope is to provide a set of generally applicable tools that may be applied to a wide range of swarms including not only single agent types, but also multiple agent types. The focus of the paper will initially rest with the development of definitions which are meant to clarify many of the often re-defined terms in swarm literature by providing mathematical definitions. Once this has been completed, the development of top-down and bottom-up equations governing swarm systems is completed. This leads to a generalized swarm condition which must be satisfied in order for the swarm to have the properties desired. The longer version of this paper [Kazadi 2005] to this paper will explore the application of these equations to two well known problems: puck collecting and puck clustering.

## 2 Swarm Definitions

One of the weaknesses in the swarm literature is the development of multiple definitions for similar concepts, each of which is open to interpretation. In any emerging field, there is significant discussion about what the "true" meaning of a concept is. In swarm intelligence and swarm engineering, this discussion has so far failed to bring universality to the definitions, and many definitions exist.

In order to both clarify our meaning and to provide an unambiguous set of definitions as a foundation for our later work, we initially examine the definitions of common terms in swarm literature. It is advantageous to provide mathematical definitions that clarify the meaning of each of the definitions. While it is not our intent that these mathematical definitions usurp the authority of definitions proposede by other swarm researchers, it is our intention to use them as a foundation for our later work.

We assume that a *system* can be thought of as a closed set of objects together with a set of consistent dynamic properties. These properties need not have closed form expressions, but we assume that they are consistent in the sense that measurements or combinations of measurements cannot produce differing numerical values for any measurable quantity. Because the system is closed, the objects are not affected by anything outside of the system. As a result, in simulations involving an outside controller of an agent in the simulation, the controller must be viewed as part of the system.

We define a *property* of the system to be a characteristic of the system that can be measured using a process that is independent of the characteristic. In what follows, we'll represent a system's property as $P_i$ where the subscript $i$ serves to identify the property. As an example, the temperature of a processor may be measured using the radiative emmissions of the processor, even though the measurement cannot affect the processor's temperature [1].

We define an *agent* [Maes 1992] to be a situated subset of the system that exhibits *autonomous control* over at least one degree of freedom in the system. Autonomous control is control which does not exhibit a direct dependence on any part of the system other than the controlling element(s); the behavior of the agent also must not be attributable to the dynamic interactive equations that define the system.

An autonomous agent is an agent that acts without the direct control of any outside influence. This means that outside of the things that it can sense, no part of the outside world affects any part of the agent's controller. While the agent can be affected by other things that it can sense, the effect of the senses is expected to be independent of the cause of the sensory input to its controller. Anything failing to meet this metric cannot be thought of as autonomous.

We may quantify this idea. Let the controller of an agent be defined by the way in which the agent responds to its memory state $M$ and its sensory state $S_s$. Then, given any outside property of the system $P_S$, it is true that if the current state of the agent is given by $S_a$ then

$$\frac{dS_a}{dt} = \frac{\partial S_a}{\partial M}\frac{dM}{dt} + \frac{\partial S_a}{\partial S_s}\frac{dS_s}{dt} + \frac{\partial S_a}{\partial P_S}\frac{dP_S}{dt}. \qquad (1)$$

That the final term is zero for all outside properties is a necessary and sufficient condition for autonomy. Now, this does not say that $\frac{dS_s}{dP_S}$ is zero. It simply means that the only way that this property may enter the controller is through the senses.

We define the *behavior of a subset of the system* to be the way in which properties of the subsystem change in time. Ie., if $P_i$ is a property, then a behavior $b_i$ is defined by

$$b_i = \frac{dP_i}{dt}. \qquad (2)$$

Behaviors often involve the interplay between more than one property. In this case, we require a formalism for describing such behaviors. Let us suppose that a system is made up of elements whose behavior is defined in terms of measurables $A = \{P_1, \ldots, P_n\}$. Then this system can have a behavior which is composed of all of the behaviors of the different measurables. That is,

$$\overrightarrow{b_A} = \left(\frac{dP_1}{dt}, \ldots, \frac{dP_n}{dt}\right). \qquad (3)$$

These properties can be most easily thought of as composite properties of many agents or objects in the system. For instance, a star has a discernable size which is defined as a combination of the positional properties of the atoms making it up. Any single element of the system would be insufficient to describe the system. Thus, the property must be described in terms of the properties of all (or at least many) of the atoms in the system.

---

[1] The radiative emmissions of the measuring device are likely to be much less important in determining the temperature of the device than internal processes. Thus, the effect of these emmissions is assumed to be negligible.

In many physical systems, there are properties that are derivations of other properties. These properties are not basic in the sense that they do not depend on dynamics of other properties. As an example, consider a point mass in our universe. We may define its position in terms of a variable $\overrightarrow{x}$. However, another property, the velocity $\overrightarrow{v}$, is a derivative property whose relationship with the basic positional property is given by

$$\overrightarrow{v} = \dot{\overrightarrow{x}}. \tag{4}$$

It is possible to measure this property of the object, and so it is indeed a property of the system, as well as a behavior of the object. This duality of behavior and property can be resolved only by noting that behaviors are linked to properties, but the behaviors can only become properties if they, in fact, can be independently measured.

Emergence has been identified by many authors in the past [Bonabeau 1999] in terms capturing the general idea that a system can have unintended global properties that are not explicitly built into its agents. The interest in swarm based systems seems to have come from this single observation. We now propose a rigorous definition of this property.

Suppose that we have a property $P_j$ that is a function of another properties and behaviors of the system. That is, suppose that

$$P_j = f\left(b_1, \ldots, b_{n_b}, P_1, \ldots, P_{j-1}, P_{j+1}, \ldots, P_{n_P}\right), \tag{5}$$

where $n_b$ is the number of systems behaviors, and $n_P$ is the number of systems properties. The number of behaviors is not necessarily equal to the number of properties of the system. The property $P_j$ is an *emergent property* of the subsystem $i$ if

$$\frac{\partial b_i}{\partial P_j} = 0. \tag{6}$$

That is, the property $P_j$ is not a factor in the defining function of behavior $b_i$ for any of the behaviors of the elements of the system. This means that the agent or agents in the system are acting independently of the property, and so the property is not a deliberate result of the design of the agent's behaviors. As a result, it cannot be viewed as part of the design of the agent(s), and so it satisfies the meaning of emergence.

Given the distinction between behaviors and properties above, we can also define *emergent behaviors* to be emergent properties that are themselves behaviors.

These definitions may be used to formally define various types of swarms of agents. Firstly, we define a *swarm of agents* to be a set of interacting agents within a system in which one agent's change in state can be perceived by at least one other agent and effects the state of the agent perceiving the change. Moreover, the subset must have the property that every agent has at least one state whose change will initiate a continual set of state changes that affects every other agent in the swarm.

Perhaps the most informative thing when working with new definitions is to examine what things do *not* satisfy the definition given. In modern times, many Westerners enjoy spending time playing video games. The past time has become so widespread that a huge industry has formed around it, and this industry includes both public and private mechanisms for playing. One of the methods of playing video games is playing on machines bought and equipped by companies devoted to providing this service. Patrons go to large rooms filled with computers, find games on the computers and play them. However, despite the fact that the patrons are all in a room, can hear various sounds produced by the computers of other patrons (as well as other patrons themselves), can see what other patrons are doing, they are acting independently. That is, their internal state is not affected by other patrons in any type of cascade effect.

However, if all the patrons begin playing the same game, interacting through the computers, the situation changes. Now, for many popular games, the patrons roam through virtual environments interacting with each other in various ways. One interaction will most certainly affect the next, with the effects rippling through the environment. Moreover, each individual will remain a member of the swarm for as long as this interaction continues, and stop when he or she gets up and leaves (or switches games). Whether or not these interactions cause any emergent behaviors is secondary, and depends on whether or not a measurable of the system is caused by the participation of the different individuals without the behaviors being affected by the property in question[2].

Let us more rigorously define a swarm. Suppose that the state of the agents is specified by a set of

---

[2] Another example of swarms and emergence typically found in the literature is traffic [Resnick, 1994]. It is understood that traffic jams are indeed emergent phenomena. However, do the agents within form a swarm? A single car's behavior (flashing brake lights, changing lane, stopping) can have a cascade effect on the cars following it. However, the effect does not travel forward. Since our requirement for a swarm was that the cascade effect affect all members of the swarm, we must conclude that traffic does not produce a swarm, though it does produce emergence, according to our definitions.

varibles $\{S_i\}$. Then the set of agents is a swarm if

$$\frac{\partial S_i|_{(t>t_0)}}{\partial S_j|_{(t=t_0)}} \neq 0\text{[3]} \tag{8}$$

$\forall i \neq j$ for times $t$ after some reference time $t_0$. That is, that the later states of agent $i$ must depend on the current state of agent $j$.

Our definition of a swarm differs from others given in the literature [Kazadi 2000] in that it does not demand emergence from the system. However, emergent swarms are also interesting, and form the basis for most of the work in swarm engineering. Thus, we define a swarm of agents as an *emergent swarm of agents with respect to property* $P_j$ if they exhibit an emergent behavior $b_{P_j}$. Note that this means that a swarm is defined only in terms of a specific property which yields the potential possibility that the group of agents is not a swarm with respect to another property $P_k$.

One of the unexpected results of this definition is that it does *not* exclude the potentiality of a centrally controlled swarm. The idea behind the swarm is that each element of the swarm is capable of initiating a cascade of state changes. How these are initiated is not important, and we can leave the possibility open that these go through a central controller, group of agents, or communication mechanism. Thus, we clarify these issues by defining a *decentralized swarm* to be a swarm that does not have a central communication or control mechanism. A *centralized swarm* is a swarm which is not decentralized.

The power of these definitions is that it is possible to test a set of agents in order to determine whether or not it is a swarm, if it is a centralized or decentralized swarm, and then whether or not it is an emergent swarm with respect to a specific property. For instance, it should be clear that a soccer team is a swarm, but it is not an emergent swarm with respect to, for instance, the team dispersion. Team members are very likely to use this information to affect their own behaviors. On the other hand, a swarm of ants is an emergent swarm with respect to food source exploitation, as it has the ability to exploit nearby food sources despite the absolute lack of knowledge on the part of the ants. This can be characterized by measuring the amount of exploitation of each food source when multiple food sources are available. Clearly this quantity is not part of the control algorithm of the agents.

## 3 Swarm engineering

Swarms are difficult to engineer primarily because groups of independent interacting agents can exhibit very complex and unexpected behaviors for a very large number of different reasons. Moreover, if the members of a group have specifications that are made independently, it is very difficult to guarantee that the specifications do not interact in this very way. Moreover, the proof of this type of interaction often requires the complete simulation of the group of agents. Finally, small perturbations to the system, which cause rather small changes in the behaviors of individual agents, can cause very large changes in the overall behavior of the system. This is, in fact, a foundational characteristic of the field of chaos [Wolfram 2002].

It is important to create a new methodology for the generation of global behavior in a way that bypasses the difficulties presented here. We seek a method that is provable in the sense that the behaviors can be understood to generate the desired global behavior. The generated behaviors have well understood tolerances for perturbations within which the desired global behavior will still occur.

As our starting point we choose the global goal. It is described in terms of a set of properties of the swarm $G = \{P_1, \ldots, P_i, \ldots, P_{n_P}\}$ and their corresponding initial and final characteristics $G^0 = \{P_1^0, \ldots, P_i^0, \ldots, P_{n_P}^0\}$ and $G^F = \{P_1^F, \ldots, P_i^F, \ldots, P_{n_P}^F\}$. The initial and final characteristics may be numerical values as in a count-based characteristic or they may be functional, as in a trajectory. They may also be sets of potential initial or final states of the two forms.

Once these initial and final conditions have been determined, it is important to specify conditions under which the final characteristics become consequences of the initial conditions and the system dynamics.

Assume that function $f$ from (5) is a differentiable function of the properties $P_i$, $i = 1, \ldots, n_p$ and the behaviors $b_i$, $i = 1, \ldots, n_b$. Then, in general case the following holds

$$b_j = \frac{dP_j}{dt} = \sum_{i=1}^{n_b} \frac{\partial P_j}{\partial b_i}\frac{db_i}{dt} + \sum_{i\neq j}^{n_p} \frac{\partial P_j}{\partial P_i} b_i. \tag{9}$$

For simplicity we assume that each property correspond with only one behavior, i.e., $n_P = n_b$, then

$$b_j = \frac{dP_j}{dt} = \sum_{i\neq j}^{n_b} \left( \frac{\partial P_j}{\partial b_i}\frac{db_i}{dt} + \frac{\partial P_j}{\partial P_i} b_i \right) + \frac{\partial P_j}{\partial b_j}\frac{db_j}{dt}. \tag{10}$$

This expresses the idea that the change in the property is a function of the connectivity between other

---

[3] This can be rigorously defined as follows:

$$\frac{\partial S_i|_{(t>t_0)}}{\partial S_j|_{(t=t_0)}} = \lim_{\delta S_j \to 0} \frac{S_i(t, S_j + \delta S_j) - S_i(t, S_j)}{\delta S_j} \neq 0 \tag{8}$$

for any time $t > t_0$.

properties of the system and the behaviors which define this property. Thus, we wish to find a set of conditions such that

$$\lim_{\tau\to\infty}\int_0^\tau \sum_{i\neq j}^{n_b}\left(\frac{\partial P_j}{\partial b_i}\frac{db_i}{dt}+\frac{\partial P_j}{\partial P_i}b_i\right)dt$$

$$+\lim_{\tau\to\infty}\int_0^\tau \frac{\partial P_j}{\partial b_j}\frac{db_j}{dt}dt+P_j^0=P_j^F. \tag{11}$$

This is the general swarm engineering condition, and must be fulfilled by the behavior and sensor sets. Behaviors of the system depend on behaviors of agents in that system. Those, in turn, depend on agents' sensors, memory state, behavioral strategy, and position. Let us assume that our swarm consists of $\{N_A\}$ agents. First, we may assume that the $l$th agent's state may be completely described by its memory state $m_s^l$, its internal state $in_s^l$, its sensor state $s_s^l$, its positional state $p_s^l$ (which expresses its position and higher derivatives of position), and its behavioral strategy $k^l$. Note that $k^l$ may be a function of time and it may be able to take on one of multiple states. Moreover, transitions may be triggered by sensor states. Then, we may express the global behavior $b_j$ as a function of a number of things. First, the coupling between a global property and an agent behavior is defined, in part, by the positional state of the agent. We define the coupling between agent $l$ and the global behavior $b_j$ by $C_{jk}^l(p_s^l, in_s^l)$. Secondly, we describe the individual behavior of the agent by $AB_{k^l}^l(m_s^l, in_s^l, s_s^l)$. Then, the overall behavior may be expressed as

$$b_j=\sum_l^{N_A} C_{jk}^l(p_s^l, in_s^l)AB_{k^l}^l(m_s^l, in_s^l, s_s^l). \tag{12}$$

[4] The trick, then, is to create behaviors that are dependant on realistic sensor states and internal states which provably satisfy equation (11). In many studies, equations $(12-14)$ are converted to average behavioral equations, greatly simplifying the required analysis.

Combining equations (11) and (12), we obtain the general combined agent-swarm equations:

$$\lim_{\tau\to\infty}\int_0^\tau \sum_{i\neq j}^{n_b}\left(\frac{\partial P_j}{\partial b_i}\frac{db_i}{dt}+\frac{\partial P_j}{\partial P_i}b_i\right)dt$$

$$+\lim_{\tau\to\infty}\int_0^\tau \frac{\partial P_j}{\partial b_j}\frac{db_j}{dt}dt+P_j^0$$

$$=\lim_{\tau\to\infty}\int_0^\tau \left[\sum_l^{N_A} C_{jk}^l(p_s^l, in_s^l)AB_{k^l}^l(m_s^l, in_s^l, s_s^l)\right]dt$$

$$+P_j^0=P_j^F. \tag{15}$$

Swarm engineering is concerned with balancing these equations linking the agent behaviors and the global desired behaviors.

## 4 Discussion

What this paper has done is generate a general way of approaching swarm design. This paper represents a step towards the purposeful art of swarm engineering, which has heretofore been without a firm theoretical basis. While many aspects of a complete theory are missing from this discussion, the use of this methodology allows one to know, before building the swarm, that the swarm will perform the global task desired [5].

What makes this method challenging to use is the fact that a suitable property must be found which leads to a useful global condition which must be satisfied. In generating these global properties, we have kept a few things in mind which would seem to be indispensible in the useful design of global properties leading to local restrictions.

1. The first step would seem to be the generation of local properties that can be accessed by the individual agents. For instance, the size of a cluster nearby an agent can be accessed using sensors that might sit atop a single agent. However, a single agent is not likely to be able to determine the dispersion of pucks.

2. The second step might be to build a global property (i.e. one that depends on measurements from more than one agent) based on the local properties that have been defined in step one. The global property should be single valued at the desired global characteristic. In generating

---

[4] In the case that there is only one behavior, this is simplified to

$$b_j=\sum_l^{N_A} C_j^l(p_s^l, in_s^l)AB^l(m_s^l, in_s^l, s_s^l). \tag{13}$$

In the additional case that there is no memory, the expression simplifies further to

$$b_j=\sum_l^{N_A} C_j^l(p_s^l, in_s^l)AB^l(in_s^l, s_s^l). \tag{14}$$

This is the equation for uniform reactive agent swarms.

[5] This, of course, assumes that the dynamics of the system are represented properly in the instantiation

such a property, it must not be possible to generate the given numerical or functional value without the desired property having the desired characteristic.

In some cases, the properties can seem contrived and somewhat unrelated to the overall system. However, the second step keeps the property grounded in the global requirement, while the first step allows the agents to be designed in such a way that they can actually be built. This methodology can be successful with potentially many global properties which may include functions, dynamics, and static end results.

The current method is neither a bottom up nor top down approach to swarm design. Rather, it is something of a middle ground between the two. While the global top-down portion of this design methodology will generate a requirement for local behaviors, it does not necessarily indicate a specific local behavior. It is still required for the behavior to be designed and then theoretically tested. This portion of the design process is bottom-up. Thus, the methodology may be characterized as a *middle-meeting* methodology.

While the current work has explored the design of swarms with specific global properties, one might expect the desired creation of swarms with multiple global goals. Such swarms will necessarily have simultaneous requirements in for the local behaviors. Of course, the design of swarms having different global properties that are in conflict with one-another is a possibility. Methods of detecting conflicting global requirements have yet to be developed but it may be possible of utilizing the current methodology to do this detection.

When the design of one global goal is not in conflict with another, and is based on a completely different local behavior, the global behaviors are parallel. In this case, it may be possible to split the swarm into castes. These castes can be made up of individual agents designed to behave in a way that meets one global goal at a time. Alternatively, if the same sensory machinery is required for both behaviors, individual agents may be recruited at different times to help carry out the differing tasks. This research may lead to new methods of developing recruitment algorithms for differing tasks at different points of task completion.

The generation of a swarm condition follows from the generation of a global property which *may not be any part of the system.* It is certainly conceivable that many properties could be generated, each leading to a swarm condition that need not be identical to the others. If this is the case, then the family of solutions could change significantly, though all of the members of the various families would be solutions to the global problem. There is, of course, no way of knowing until the analysis is completed.

This paper has not applied this methodology to any specific problem; a longer version of this paper [Kazadi 2005] includes applications of the theory.

## References

[Bonabeau 1999] E. Bonabeau, M. Dorigo, and G. Theraulaz. **Swarm Intelligence**. New York: Oxford University Press, 1999.

[Goldberg 2000] D. Goldberg and M. Matarić, *Interference as a Tool for Designing and Evaluating Multi-Robot Controllers*, **Proceedings, AAAI-97**, Providence, Rhode Island, July 27-31, 1997.

[Jones 2004] C. Jones and M. Matarić. *Automatic synthesis of communication-based coordinated multi-robot systems.* **IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, 381-387, Sendai, Japan, Sept. 2004.

[Kazadi 2000] S. Kazadi. **Swarm Engineering**. Ph.D. Thesis, California Institute of Technology, 2000.

[Kazadi 2004] S. Kazadi, M. Chung, B. Lee, R. Cho. *On the dynamics of clustering systems.* **Robotics and Autonomous Systemsr**, 46 (1), 1-27, 2004.

[Kazadi 2005] S. Kazadi. *Swarm engineering paradigm and application.* Submitted, 2005.

[Lerman 2001] K. Lerman and A. Galstyan. *A general methodology for mathematical analysis of multi-agent systems.* USC Information Sciences Technical Report ISI-TR-529, 2001.

[Maes 1992] P. Maes. *How to Do the Right Thing.* **Connection Science Journal**, 1(3), 1989

[Maris and Boekhorst 1995] *Exploiting physical constraints: heap formation through behavioral error in a group of robots.*

[Spears 2004] W. Spears, D. Spears, J. Hamann, and R. Heil. *Distributed, physics-based control of swarms of vehicles.* **Autonomous Robots**, 17(2-3), 2004.

[Wolfram 2002] **A New Kind of Science**. Champaigne: Wolfram Press, 2002.