

Evolutionary Transitions as a Metaphor for Evolutionary Optimisation

Anne Defaweux¹, Tom Lenaerts², and Jano van Hemert³

¹ COMO, Vrije Universiteit Brussel, Brussels, Belgium

² IRIDIA, CP 194/6, Université Libre de Bruxelles, Brussels, Belgium

³ Center for Emergent Computing, Napier University,
Edinburgh, United Kingdom

adefaweu@vub.ac.be, tlenaert@ulb.ac.be

Abstract. This paper proposes a computational model for solving optimisation problems that mimics the principle of evolutionary transitions in individual complexity. More specifically it incorporates mechanisms for the emergence of increasingly complex individuals from the interaction of more simple ones. The biological principles for transition are outlined and mapped onto an evolutionary computation context. The class of binary constraint satisfaction problems is used to illustrate the transition mechanism.

1 Introduction

From biological literature one can learn that life is organised in a hierarchical fashion and that transitions in complexity have occurred linking the different levels of this hierarchy. Typical examples in this context are the transitions from genes to simple cells, from single cells to multi-cellular organisms or from single organisms to social systems [1,2]. It has been argued that these transitions in the complexity of the evolving individuals share two common themes: (1) the emergence of cooperation among individuals at a lower level in the hierarchy into the functioning of a new higher level unit and (2) the regulation of conflict among these lower level units.

In this article, the metaphor defined by transitions in biological complexity is used to construct an artificial evolutionary system which can be used in the context of optimisation and learning. The central problem we investigate is how a system can be designed that captures the two themes of cooperation and mediation proposed by Michod [2] into a suitable algorithm. Hence, this article will discuss a mapping between the abstract scheme that captures the common structure of evolutionary transitions and an artificial evolutionary model that can serve as an alternative for the simple genetic algorithm (GA).

We focus here on system that provides transitions in the context of the solution complexity for the class of binary constraint satisfaction problems (BINCSPP). This example from optimisation was chosen for four reasons: (1) we were interested in a problem where solutions can be modelled by the aggregation of lower level (partial) solutions, (2) cooperative interactions between the

partial solutions can be defined in a natural way, (3) when the interactions benefit the partners, the new unit of selection that emerges at the higher-level can still be interpreted using semantics defined by the problem under observation, and (4) previous studies enable us to create problems with a controlled level of difficulty [3]. Consequently from the first three reasons, the emergent unit still has some meaningful functionality in the context of the problem.

The difference with the GA approach to evolution is that solutions are variable length representations which increase in complexity, individuals use only replication and mutation and are placed in an interactive framework which supports collaborative behaviour. Consequently, the proposed model is related to messy Genetic Algorithms (mGA) [4] and the Compositional Evolution model [5,6]. For details on the technical differences, we refer to [7]. Conceptually, the difference is in the metaphor used to construct the model. Here, as mentioned earlier, the transition perspective focuses on the one defined in [2].

In the next section, the class of optimisation problems for which the transition model will be defined is explained. Given this problem context, a mapping is examined between the transition cycle and the proposed evolutionary optimisation system. Afterwards an illustrative experiment is performed to demonstrate the increase of complexity and its effect on the fitness.

2 Optimisation Context for Transition Study

Constraint Satisfaction Problems (CSP) [8] form a NP-complete problem class where, on the one hand, one has a set of variables X associated with possible domain values D and, on the other hand, a set of constraints C defined on this set of variables, which prohibits combinations of assignments to occur. The problem consists in finding an assignment to the whole set of variables from the associated domain values so that all constraints are satisfied. If this proves to be impossible then the corresponding problem is said to be unsolvable.

A variant of this problem is BINCSPP, where each constraint is defined on at most two variables. This forms no restriction on the general form of CSP as every CSP can be rewritten into a BINCSPP and vice versa [9].

Let us take as an illustration the following BINCSPP: consider a set of six variables: $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ all taking values in $D = \{1, 2, 3\}$. We consider the following set of constraints:

$$C = \{(x_1 \neq x_2), (x_2 \neq x_3), (x_3 \neq x_1), \\ (x_4 \neq x_5), (x_5 \neq x_6), (x_6 \neq x_4), \\ (x_1 = x_4), (x_2 = x_5), (x_3 = x_6)\} \quad (1)$$

This setup of constraints consists of nine binary constraints. Each binary constraint defines a relation between two variables of X . Also, for each pair of variables, only one binary constraint may be defined.

The problem involves finding the correct assignment for the variables so that all these constraints are satisfied. We denote the assignment of one variable $x_i \in X$ with value $d \in D$ by $\langle d, i \rangle$ where i is the index of the variable we consider.

Using this notation, we represent the simultaneous assignment of variables x_1 , x_2 and x_4 with respective values v_1 , v_2 and v_4 as

$$(\langle v_1, 1 \rangle, \langle v_2, 2 \rangle, \langle v_4, 4 \rangle) \quad (2)$$

A *solution* for a BINCSP problem consists in an assignment of variables from X to values of D . We use randomly generated problem instances of BINCSP. The RandomCSP package [10] is used to generate the suite of test problem instances [11]. To scale the difficulty of the problem instances, these CSP are generated according to two parameters. For more details see [11,7]. An important property that was observed is that for certain problem instances there is more structure in the search space than others. Whenever structure is present the algorithm described later can exploit it.

3 Evolutionary Transition in BINCSP Solutions

As mentioned in the introduction, all transitions in nature share two common themes: cooperation and conflict mediation among the lower-level individuals. These themes are captured in the transition cycle visualised in Figure 1. One can observe four phases in this cycle and these phases need to be captured by the proposed algorithm.

A system that uses the metaphor visualised in the figure, should be able to apply it iteratively. In other words, repeated phases of cooperation and mediation between ever increasing levels will produce more and more complex organisms which try to survive in their selective environment. In the following sections a mapping will be defined of an optimisation process onto the cycle. For the rest of the discussion, it is assumed that an evolutionary system is present that simulates the process of differential survival and reproduction of the partial and complete solutions for a particular BINCSP problem.

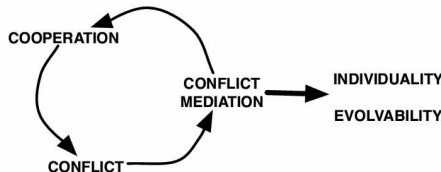


Fig. 1. Transition Cycle [2]; Every transition starts at a certain level of complexity. At this level cooperation needs to emerge since it exchanges fitness at the lower level with fitness at the higher level. Yet conflict remains. Defection among the lower-level units can lead to the destruction of the cooperative group. These conflicts need to be mediated and this will lead to a new level of individuality with its own heritable variations that evolve and diversify.

3.1 Representation of Lower-Level Units

At the lowest (initial) level, the system consists of a population P containing N individuals where each individual i is represented by a set of variables $S_i \subseteq X$ for all $i \in \{1, 2, \dots, N\}$. Here it is assumed that, at the lowest level, the size of this set S is 2. Hence, the initial population contains only *partial* solutions which can solve one of the binary constraints in the set C (see Equation (1)). An individual which contains a value for all variables in X is referred to as a *complete* or *fully qualifying* solution. Hence, *complexity* in the current system refers to the number of variables present in an individual i.e. individuals of maximum complexity are complete solutions.

A partial solution s that only defines values for x_1 and x_2 is for example:

$$(\langle 1, 1 \rangle, \langle 3, 2 \rangle). \tag{3}$$

(3) is called the *genotype* of the solution. The selective system will operate on the quality of the genotype in solving the constraints listed in the set C .

3.2 Cooperation Between Lower-Level Units

Interactions between the partial individuals is done between pairs of individuals that are randomly selected from the population P . The experiments discussed here will not consider larger groups. This assumption is removed in some ongoing experiments, yet they will not be reported here.

Interaction between partial solutions is defined in the following manner. Let solution s defined by (3) interact with a symbiotic partner sp defined by $(\langle 3, 1 \rangle, \langle 2, 3 \rangle)$. This interacting partner is referred to as the symbiotic partner of (3) and we denote the relation by:

$$(\langle 1, 1 \rangle, \langle 3, 2 \rangle) \leftrightarrow (\langle 3, 1 \rangle, \langle 2, 3 \rangle). \tag{4}$$

We simulate interaction between 2 partial solutions by sharing the information contained in their genotypes. The outcome of the information sharing between a solution and its symbiotic partner is called the *phenotype* of the solution. For example, the phenotype of (4) is simply obtained by combining all information present in both genotypes:

$$\left\langle \begin{pmatrix} 1 \\ 3 \end{pmatrix}, 1, \langle 3, 2 \rangle, \langle 2, 3 \rangle \right\rangle \tag{5}$$

Important to notice here is that the genetic information of both individuals is not changed. The heritable capacity of both s and sp remains at the level of the simple units.

The situations one can have when combining s and sp correspond to the different general forms of symbiosis: parasitism (P), mutualism (M), commensalism (C) and amensalism (A). In the case of parasitism, as shown in Table 1, the association is disadvantageous for one of the partners and beneficial to the

other one. The outcome of the interaction between s and sp is parasitic if s solves one of the constraints of C correctly ($A(s) = \text{high}$) and sp does not ($A(sp) = \text{low}$). The information sharing will in that case benefit sp since it increases its adaptiveness and it decreases the adaptiveness of s . The relation between s and sp is mutualistic if both partners gain something from the relation ($A(s, sp) = \text{high}$). In the table, both individuals have low adaptiveness but when the two variables x_2 and x_3 are combined, i.e. $(\langle 3, 2 \rangle, \langle 2, 3 \rangle)$, their adaptiveness increases. Commensalism occurs when the adaptiveness of one of the partners does not change due to the information sharing. In the table, an example is shown where both individuals have a value for variable x_1 i.e. the values 1 and 3. Now if the assignment $x_1 = 1$ resolves one of the constraints and it is assumed that the value is selected by both partners then only sp benefits from the relation and things do not change for s . A similar reasoning can be followed for amensalism. In that case the association is disadvantageous for one of the partners. Yet then instead of choosing the best value for x_1 , the worst one is selected.

3.3 Conflict Mediation

Although cooperative behaviour produces better results in the long term, short term considerations will lead to defecting behaviour. An important choice made by individuals in a transition model is whether they will share the information or not. In order to have transitions in complexity, mechanisms should be put into place which encourage the evolution of information sharing behaviour. In the current model, it is assumed that individuals want to collaborate. In other words they are all cooperative. In further experiments this assumption is relaxed. This simplification was made to examine whether cooperative partially defined units can actually lead to fully qualifying solutions for BINCSP problems. In general, principles from multilevel selection are incorporated to model the evolution of cooperative interactions between pairs (or between members of more complex groups) [12]. For now, we focus on another conflict issue.

Next to the choice of collaborating or not, other conflicts can occur. As shown in (5) partners can have different values for the same variables. These

Table 1. Some examples of the different forms of symbiosis and their relation to the BINCSP problem. $A(s)$ and $A(sp)$ evaluate the adaptiveness of both individuals in their personal relation to the problem. $A(s, sp)$ refers the effects of the information sharing on the adaptiveness of both individuals.

	s	$A(s)$	sp	$A(sp)$	symbiosis	$A(s, sp)$
P	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle)$	high	$(\langle 3, 3 \rangle, \langle 2, 4 \rangle)$	low	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle, \langle 3, 3 \rangle, \langle 2, 4 \rangle)$	(low,high)
M	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle)$	low	$(\langle 3, 1 \rangle, \langle 2, 3 \rangle)$	low	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle, \langle 3, 1 \rangle, \langle 2, 3 \rangle)$	(high,high)
C	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle)$	high	$(\langle 3, 1 \rangle, \langle 2, 3 \rangle)$	low	$(\langle \begin{pmatrix} 1 \\ 3 \end{pmatrix}, 1 \rangle, \langle 3, 2 \rangle, \langle 2, 3 \rangle)$	(high,high)
A	$(\langle 1, 1 \rangle, \langle 3, 2 \rangle)$	high	$(\langle 3, 1 \rangle, \langle 2, 3 \rangle)$	low	$(\langle \begin{pmatrix} 1 \\ 3 \end{pmatrix}, 1 \rangle, \langle 3, 2 \rangle, \langle 2, 3 \rangle)$	(low,low)

problems with conflicting values are resolved by selecting randomly one of the possible choices. Hence, the symbiotic behaviour can correspond to any of those described in Table 1.

In our example, a conflict needs to be resolved for variable x_1 . We can choose between the values 1 and 3. A possible conflict resolution in this case would be:

$$\langle 1, 1 \rangle, \langle 3, 2 \rangle, \langle 2, 3 \rangle \quad (6)$$

(6) is called the *induced phenotype* of the partial solution (3). This phenotype is used for evaluation and the result of the evaluation is assigned to the genotype s i.e. (3). We denote the phenotype of a solution s interacting with sp by: $\phi(s, sp)$.

The phenotype assigned to the symbiotic partner sp is obtained in the same way. Yet, the policy about conflicting values may yield another representation than the one we obtained for the initial partial solution s . This asymmetry between the phenotype of a solution and the phenotype of its symbiotic partner increases the exploration possibilities of the evolutionary process. Note that the conflict mediation strategy adopted for this particular test case avoids the system to build greater genotype than the maximum size expected for a genotype. Hence, the problem related to ever growing genotypes which is a classical issue in variable length representation does not occur here.

3.4 Intermezzo: Evaluation of Genotypes

Here two types of functions are considered. One function to determine the success of the solution in terms of the complete constraint set ($f(s)$) and another function to determine how good it scores relative to the constraints it covers ($f_{cov}(s)$).

Assume that $c_k(p)$ is the outcome of evaluating phenotype p with constraints k , we say that p covers c_k if p contains assignments for all variables contained in c_k , furthermore, p satisfies c_k if the assignment values in p do not violate the constraints defined by c_k .

$$c_k(p) = \begin{cases} 1 & \text{if } p \text{ covers } c_k \wedge p \text{ satisfies } c_k \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Given this, the classical evaluation of the solution described by (3) and denoted by s working with a symbiotic partner sp is given by:

$$f(s) = \frac{1}{|C|} \sum_{k \in C} c_k(\phi(s, sp)) \quad (8)$$

where C is the constraints set, $|C|$ the size of the constraints set and $\phi(s, sp)$ the induced phenotype of s when sharing information with its symbiotic partner.

It does however not give any indication of the quality of the partially defined assignments *relatively to the constraints it covers*. To see whether an association works fine or not, a restricted fitness measure is define that only considers the constraints covered by the phenotype of the solution.

Assume that $cov(s, C)$ is the set of constraints covered by s , the covering fitness measure is given by:

$$f_{cov}(s) = \frac{1}{|cov(\phi(s, sp), C)|} \sum_{k \in cov(\phi(s, sp), C)} c_k(\phi(s, sp)) \quad (9)$$

We use the first measure (8) to guide the evolutionary process (selection). The second measure is used by the evolutionary observer to decide whether a transition should be performed. The idea of introducing a mechanistic observer to decide when a transition occurs corresponds to the work in [13]. Important to remember is that although the fitness is determined using the induced phenotype the fitness value is assigned to the genotype. Hence the process of differential survival and reproduction operates at the genotype level and not at the level of the induced phenotype.

3.5 Higher-Level Individuality and Evolvability

As was assumed in the beginning of this section, solutions (genotypes) are selected according to their fitness described by (8). When a solution is selected, it will replicate into a new solution. There is a certain probability that this replication process has errors and in this way mutants can emerge.

The symbiosis between s and sp also has some consequences for the reproductive process. In certain circumstances beneficial symbiotic relations will be replicated as a whole. When the symbiotic partner is replicated as well, the symbiotic link, that is, their interaction scheme will be inherited in the process. The underlying idea is that (possibly) good working units can survive over more than one generation. The idea of performing this replication in group is based on our previous work in the context of multi-level selection [12]. For now it is assumed that decision to replicate the group is decided randomly using a probability q (here, $q = 0.5$). More elaborate methods based on the type of interaction can be used. Note that there is still a probability $(1 - q)$ of individual replication.

This replication of the both genotypes is a *first step* toward a new higher-level entity of selection. Although simple lower-level entities can sometimes replicate in group they still have the possibility of spreading their own genetic material (probability $(1 - q)$). In the *second step*, both partners give up their individual replication process in favour of a group replication process. At that point, the transition has occurred since replication becomes now the responsibility of the higher-level structure. To make this final step the function $f_{cov}(s)$, defined in the previous section, is used (see Equation (9)). When the induced phenotype happens to solve the sub-problem defined by the covering set of constraints, i.e. when f_{cov} has reached a certain threshold value (here we selected $f_{cov}(s) = 1$), a new more complex individual is created whose genotype corresponds to the induced phenotype of the previous symbiotic relation.

Let us take the example solution previously discussed. The solution described by (3) with the phenotype given in (6) has a classical fitness value of $f(s) = 0.33$. The measure of the fitness restricted to the covering constraints set was

$f_{cov}(s) = 1$. In this case, if the solution is selected, the system creates a higher level unit combining the genetic information of both partners. This means that the expression of the genotype of the new of unit is:

$$((1, 1), \langle 3, 2 \rangle, \langle 2, 3 \rangle) \quad (10)$$

This defines the transition step: Solutions are incrementally grown according to their success in solving the sub-problem they are defined for.

The information sharing strategy we adopted allows information to be exchanged when values are conflicting on certain allele (such as this was the case in example (5) on variable 1). This mimics a cross-over operation as full grown solution are interacting. When partial solutions interact, the exchange can only happen on conflicting parts of the genotype, yet, preserving the non conflicting parts of the genotypes. In this case, we can consider this as a preserving cross-over operation that avoids recombining good part of the solutions. In this way, the model uses the notion of compositional evolution as discussed in [5,6].

4 Identifying Transitions in Complexity

In this section, we will illustrate the transition process by analysing the outcome of a simple simulation of this model for random generated BINCSP instances of 15 variables each taking values in a domain of size 15 which are made easy or difficult by tuning parameters such as the density of the constraints network p_1 or the average tightness within the constraints \bar{p}_2 [11]. We propose to discuss the results for two different setups of these parameters which yield respectively a relatively easy ($\bar{p}_2 = 0.3$ and $p_1 = 0.9$) and relatively difficult ($\bar{p}_2 = 0.5$ and $p_1 = 0.5$) BINCSP instance. For each setup, we solved 25 instances and performed 10 runs for each instance. The 250 runs were then analysed by looking at the fitness relatively to the genotype size and the evolution of the size and fitness over time. Increased size of the genotype reflects successive transitions from simple units starting from length 2 up to a complex units that solves greater number of constraints.

In Figure 2, we plotted for each setup the fitness and genotype size dynamics for an isolated run, the average on all the runs of the fitness with respect to the size and the average on all the runs of the duration a genotype remains at a certain level before performing a transition.

In the first row of Figure 2 (the genotype and induced phenotype size have been rescaled to $[0, 1.0]$ for illustration purposes), a close relationship between the trend of the genotype size and the fitness trend can be observed. We can conclude from this that transitions are needed to allow the fitness to reach higher levels. For difficult problem instances, once the genotype size and the resulting phenotype size are fixed, we can observe that the fitness value still slowly improves over time. This slow improvement illustrates a phase of conflict mediation where the partial solution and its symbiotic partner try to reduce the set of variables which yield conflicts. For easy problem instances, Fitness is closely related to the genotype size, yet, we can observe that a good working symbiotic

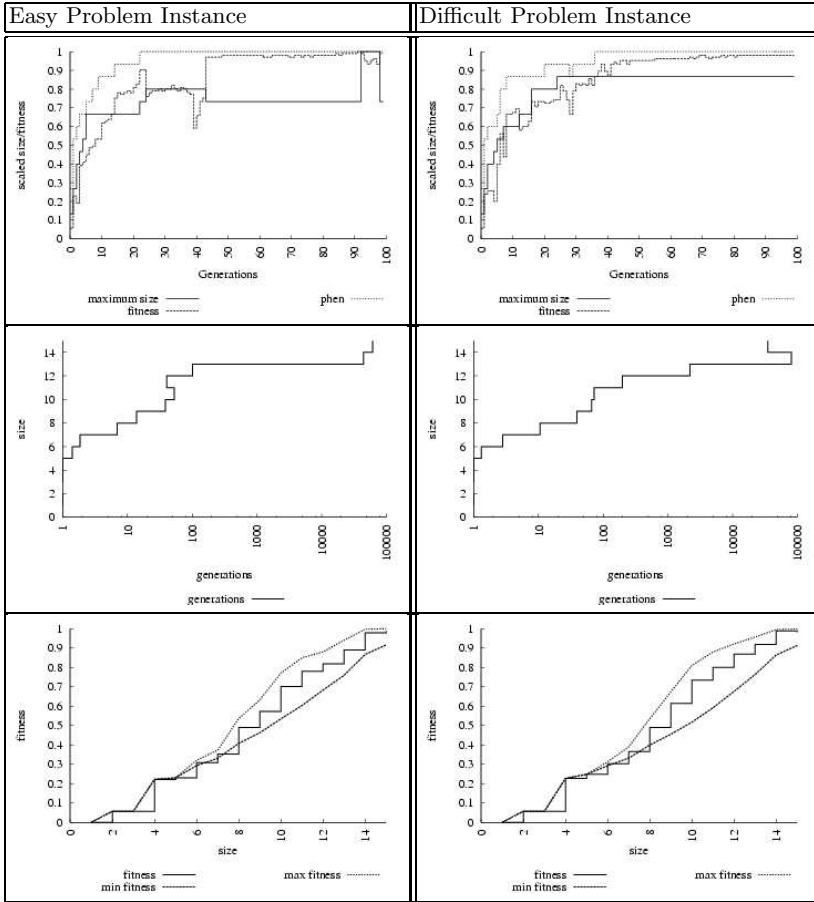


Fig. 2. The left column gives the results for an easy test case instance while the right column gives the same result for a difficult test case instance. On the first row, the evolution of the fitness relatively to the scaled genotype and phenotype size for one simulation run is plotted. On the second row, the average required time for the next transition to occur and on the last row, the fitness that corresponds to each genotype size.

relation has been discovered by the process (which can be seen on the graph by a relatively small genotype size for which the corresponding phenotype defines a complete solution). This good working collaboration is sustained for a while before the conflicts among the symbiotic partners are resolved and a transition can occur.

In the second row, which refers to the time required for different genotype sizes to evolve toward a new level of complexity, we see that the time increases as the genotypes become more complex. In other words, the conflict mediation becomes more difficult as the interacting units grow in complexity. The tran-

sitions which occurred at fast speed in the beginning require more time as the evolutionary process goes on and the conflicts to be resolved (the variables which share different values in the solution and its symbiotic partner) increase. We observe that this phenomenon is independent from the hardness of the problem instance as the increased number of generations required to master a new level of complexity is observed in both cases. Note also that this increase in time to move between complexity levels has also been observed in nature [1].

Finally, a look at the fitness relative to the genotype size (averaged over all runs) confirms the first observation that the increasing complexity at the genotype level results in an increase of the functionality of this genotype. In other words, the system requires transitions to attain the level of complexity specified by the problem instance.

5 Conclusions

In this paper, we addressed the issue of the emergence of complexity in evolutionary optimisation algorithm. Inspiration was found in the theories concerning evolutionary transitions observed in Biology. These theories propose a generalised explanation for the mechanism by which interacting lower level units can produce new higher level ones. The proposed Transition algorithm uses symbiosis as the basic ingredient for the system to work. To illustrate this model, we applied it to BINCSP and showed within this context how the mechanism of transition worked. The algorithm was also compared thoroughly to other evolutionary approaches for solving BINCSP [7]. These experiments showed the great promise for the discussed technique.

References

1. Maynard Smith, J., Szathmáry, E.: The major transitions in evolution. Oxford University Press (1995)
2. Michod, R.: Darwinian Dynamics: Evolutionary transitions in Fitness and Individuality. Princeton University Press (1999)
3. Palmer, E.M.: Graphical Evolution. John-Wiley & Sons, New York (1985)
4. Goldberg, D., Korb, B., Deb, K.: Messy genetic algorithms: motivation, analysis, and first results. *Complex Systems* **3** (1989) 493–530
5. Watson, R.A., Pollack, J.B.: Symbiotic combination as an alternative to sexual recombination in genetic algorithms. In: *Parallel Problem Solving from Nature (PPSNVI)*, Springer Verlag, Lecture Notes in Computer Science 1917 (2000)
6. Watson, R.A., Pollack, J.B.: A computational model of symbiotic composition in evolutionary transitions. *Biosystems Special Issue on Evolvability* **69/2-3** (2002) 187–209
7. Defaweux, A., Lenaerts, T., van Hemert, J., Parent, J.: Transition models as an incremental approach for problem solving in evolutionary algorithms. In: *the Genetic and Evolutionary Computation Conference*, to appear (2005)
8. Tsang, E.: *Foundations of Constraint Satisfaction*. Academic Press Limited (1993)

9. Rossi, F., Dhar, V.: On the equivalence of constraint satisfaction problems. In Aiello, L.C., ed.: ECAI'90: Proceedings of the 9th European Conference on Artificial Intelligence, Stockholm, Pitman (1990) 550–556
10. van Hemert, J.: RandomCSP (2004) Freely available from <http://freshmeat.net/projects/randomcsp/>.
11. van Hemert, J.: Application of Evolutionary Computation to Constraints Satisfaction and Data Mining. PhD thesis, Universiteit Leiden (2002)
12. Lenaerts, T., Defaweux, A., van Remortel, P., Nowé, A.: Evolutionary game dynamics of intrademic multilevel selection. Technical Report TR/IRIDIA/2005-07, IRIDIA (2005)
13. Bersini, H.: Whatever emerges should be intrinsically useful. In: Proceedings of the ninth international conference on artificial life, MIT press (2004) 226–231