# Elicitation Strategies for Fuzzy Constraint Problems with Missing Preferences: an Experimental Study

M. Gelain\*, M.S. Pini\*, F. Rossi\*, K. Venable\* and T. Walsh\*\*

\* University of Padova, Italy, E-mail: {mgelain,mpini,frossi,kvenable}@math.unipd.it \*\* NICTA and UNSW Sydney, Australia, Email: Toby.Walsh@nicta.com.au

#### Abstract

Fuzzy constraints are a popular approach to handle preferences and over-constrained problems. We consider here situations where some of the preferences may be missing. This models, for example, settings where agents are distributed, or have privacy issues, or where there is an ongoing preference elicitation process. We study how to find a solution which is optimal irrespective of the missing preferences, eliciting preferences from the user if necessary. Our goal is to ask the user as little as possible. To solve this task, we define a combined solving and preference elicitation scheme with a large number of different instantiations which we test on randomly generated problems. Our experimental results show that some of the algorithms are very good at finding a necessarily optimal solution while asking only a very small fraction of the missing preferences to the user. We also test the algorithms on hard constraint problems with possibly missing constraints. The aim now is to find feasible solutions irrespective of the missing constraints.

#### Introduction

Constraint programming is a powerful paradigm for solving scheduling, planning, and resource allocation problems. A problem is represented by a set of variables, each with a domain of values, and a set of constraints. A solution is an assignment of values to the variables which satisfies all constraints and which optionally maximizes/minimizes an objective function. It is usually assumed that the data (variables, domains, constraints) is completely known before solving starts. This is often unrealistic. In web applications and multi-agent systems, the data is frequently only partially known and may be added to a later date by, for example, elicitation. Data may also come from different sources at different times. In multi-agent systems, agents may release data reluctantly due to privacy concerns.

Incomplete soft constraint problems can model such situations by allowing some of the preferences to be missing. An algorithm has been proposed and tested to solve such incomplete problems (Gelain et al. 2007). The goal is to find a solution that is guaranteed to be optimal irrespective of the missing preferences, eliciting preferences if necessary until such a solution exists. Two notions of optimal solution are considered: *possibly optimal* solutions are assignments that are optimal in *at least one way* of revealing the unspecified preferences, while *necessarily optimal* solutions are assign-

ments that are optimal in *all ways* that the unspecified preferences can be revealed. The set of possibly optimal solutions is never empty, while the set of necessarily optimal solutions can be empty.

If there is no necessarily optimal solution, the algorithm proposed in (Gelain et al. 2007) uses branch and bound to find a "promising solution" (specifically, a complete assignment of the best possible completion of the current problem) and elicits the missing preferences related to this assignment. This process is repeated till there is a necessarily optimal solution. Although this algorithm behaves reasonably well, it make some specific choices about solving and preference elicitation that may not be optimal in practice, as we shall see in this paper. For example, the algorithm only elicits missing preferences after running branch and bound to exhaustion. As a second example, the algorithm elicits all missing preferences related to the candidate solution. Many other strategies are possible. We might elicit preferences at the end of every complete branch, or even at every node in the search tree. When choosing the value to assign to a variable, we might ask the user (who knows the missing preferences) for help. Finally, we might not elicit all the missing preferences related to the current candidate solution. For example, we might just ask the user for the worst preference among the missing ones.

In this paper we consider a general algorithm scheme which generalizes that proposed in (Gelain et al. 2007). It is based on three parameters: *what* to elicit, *when* to elicit it, and *who* chooses the value to be assigned to the next variable. We test 16 different instances of the scheme on randomly generated fuzzy constraint problems. We show that some algorithms are very good at finding necessarily optimal solution without eliciting too many preferences. We also test the algorithms on problems with hard constraints. Finally, we consider problems with fuzzy temporal constraints, where problems have more specific structure.

In our experiments, besides computing the percentage of elicited preferences, we also compute the user's effort when we ask for their help in the search process. For instance, when we ask the user for the worst preference, we are asking them to do some work for us. This effort is therefore also an important measure. Our results show that the choice of preference elicitation strategy is crucial for the performance of the solver. While the best algorithms need to elicit as

little as 10% of the missing preferences, the worst one need much more. The performance of the best algorithms also shows that we only need to ask the user a very small amount of additional information to be able to solve problems with missing data.

Several other approaches have addressed similar issues. For example, open CSPs (Faltings and Macho-Gonzalez 2002; 2005) and interactive CSPs (Lamma et al. 1999) work with domains that can be partially specified. As a second example, in dynamic CSPs (Dechter and Dechter 1988) variables, domains, and constraints may change over time. However, the incompleteness considered in (Faltings and Macho-Gonzalez 2005; 2003) is on domain values as well as on their preferences. We assume instead, as in (Gelain et al. 2007), that all values are given at the beginning, and that only some preferences are missing. Because of this assumption, we don't need to elicit preference values in order, as in (Faltings and Macho-Gonzalez 2005). In (Braziunas and Boutilier 2006) preference elicitation is performed in the generalized additive independence model, that compactly represents both linear utility functions and graphical models like UCP-nets (Boutilier, Bacchus, and Brafman 2001). Our paper, differently from (Braziunas and Boutilier 2006), is based on the soft constraint formalism, and thus it does not force the user to express his preferences via utility functions. However, we are not able to handle also qualitative preferences.

## **Background**

Incomplete Soft Constraints problems (ISCSPs) (Gelain et al. 2007) extend Soft Constraint Problems (SCSPs) (Bistarelli, Montanari, and Rossi 1997) to deal with partial information. We will focus on a specific instance of this framework in which the soft constraints are fuzzy. Given a set of variables V with finite domain D, an  $incomplete\ fuzzy$  constraint is a pair  $\langle idef, con \rangle$  where  $con \subseteq V$  is the scope of the constraint and  $idef: D^{|con|} \longrightarrow [0,1] \cup \{?\}$  is the preference function of the constraint associating to each tuple of assignments to the variables in con either a preference value ranging between 0 and 1, or ?. All tuples mapped into ? by idef are called  $incomplete\ tuples$ , meaning that their preference is unspecified. A fuzzy constraint is an incomplete fuzzy constraint with no incomplete tuples.

An incomplete fuzzy constraint problem (IFCSP) is a pair  $\langle C, V, D \rangle$  where C is a set of incomplete fuzzy constraints over the variables in V with domain D. Given an IFCSP P, IT(P) denotes the set of all incomplete tuples in P. When there are no incomplete tuples, we will denote a fuzzy constraint problem by FSCP.

Given an IFCSP P, a completion of P is an IFCSP P' obtained from P by associating to each incomplete tuple in every constraint an element in [0,1]. A completion is partial if some preference remains unspecified. C(P) denotes the set of all possible completions of P and PC(P) denotes the set of all its partial completions.

Given an assignment s to all the variables of an IFCSP P, pref(P,s) is the preference of s in P, defined as  $pref(P,s) = min_{<idef,con> \in C|idef(s_{\downarrow con}) \neq ?idef(s_{\downarrow con})}$ . It is obtained by taking the minimum among the known pref-

erences associated to the projections of the assignment, that is, of the appropriated sub-tuples in the constraints.

In the fuzzy context, a complete assignment of values to all the variables is an optimal solution if its preference is maximal. The optimality notion of FCSPs is generalized to IFCSPs via the notions of necessarily and possibly optimal solutions, that is, complete assignments which are maximal in all or some completions. Given an IFCSP P, we denote by NOS(P) (resp., POS(P)) the set of necessarily (resp., possibly) optimal solutions of P. Notice that  $NOS(P) \subseteq POS(P)$ . Moreover, while POS(P) is never empty, NOS(P) may be empty. In particular, NOS(P) is empty whenever the revealed preferences do not fix the relationship between one assignment and all others.

In (Gelain et al. 2007) an algorithm is proposed to find a necessarily optimal solution of an IFCSP based on a characterization of NOS(P) and POS(P). This characterization uses the preferences of the optimal solutions of two special completions of P, namely the **0**-completion of P, denoted by  $P_0$ , obtained from P by associating preference 0 to each tuple of IT(P), and the 1-completion of P, denoted by  $P_1$ , obtained from P by associating preference 1 to each tuple of IT(P). Notice that, by monotonicity of min, we have that  $pref_0 \leq pref_1$ . When  $pref_0 = pref_1$ ,  $NOS(P) = Opt(P_0)$ ; thus, any optimal solution of  $P_0$  is a necessary optimal solution. Otherwise, NOS(P) is empty and POS(P) is a set of solutions with preference between  $pref_0$  and  $pref_1$  in  $P_1$ . The algorithm proposed in (Gelain et al. 2007) finds a necessarily optimal solution of the given IFCSP by interleaving the computation of  $pref_0$  and  $pref_1$ with preference elicitation steps, until the two values coincide. Moreover, the preference elicitation is guided by the fact that only solutions in POS(P) can become necessarily optimal. Thus, the algorithm only elicits preferences related to optimal solutions of  $P_1$ .

## A general solver scheme

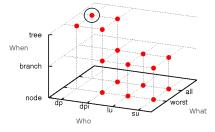
We now propose a more general schema for solving IFCSPs based on interleaving branch and bound (BB) search with elicitation. This schema generalizes the concrete solver presented in (Gelain et al. 2007), but has several other instantiations that we will consider and compare experimentally in this paper. The scheme uses branch and bound. This considers the variables in some order, choosing a value for each variable, and pruning branches based on an upper bound (assuming the goal is to maximize) on the preference value of any completion of the current partial assignment. To deal with missing preferences, branch and bound is applied to both the 0-completion and the 1-completion of the problem. If they have the same solution, this is a necessarily optimal solution and we can stop. If not, we elicit some of the missing preferences and continue branch and bound on the new 1-completion.

Preferences can be elicited after each run of branch and bound (as in (Gelain et al. 2007)) or during a BB run while preserving the correctness of the approach. The algorithm schema we propose is based on the following parameters:

 Who chooses the value of a variable: the algorithm can choose the values in decreasing order either w.r.t. their preference values in the 1-completion (Who=dp) or in the 0-completion (Who=dpi). Otherwise, the user can suggest this choice. To do this, he can consider all the preferences (revealed or not) for the values of the current variable (*lazy user*, Who=lu for short); or he considers also the preference values in constraints between this variable and the past variables in the search order (*smart user*, Who=su for short).

- 2. **What** is elicited: we can elicit the preferences of all the incomplete tuples of the current assignment (What=all) or only the worst preference in the current assignment, if it is worse than the known ones (What = worst);
- 3. **When** elicitation takes place: we can elicit preferences at the end of the branch and bound search (When=tree), or during the search, when we have a complete assignment to all variables (When =branch) or whenever a new value is assigned to a variable (When = node).

By choosing a value for each of the three above parameters in a consistent way, we obtain 16 different algorithms, as summarized in the figure below, where the circled instance is the concrete solver used in (Gelain et al. 2007).



```
\begin{split} & \mathsf{IFCSP\text{-}SCHEME}(P, Who, What, When) \\ & Q \leftarrow P_0 \\ & s_{max}, pref_{max} \leftarrow BB(P_0, -) \\ & Q', s_1, pref_1 \leftarrow BBE(P, 0, Who, What, When, s_{max}, pref_{max}) \\ & \mathsf{If} \ (s_1 \neq nil) \\ & s_{max} \leftarrow s_1, pref_{max} \leftarrow pref_1, Q \leftarrow Q' \\ & \mathsf{Return} \ Q, s_{max}, pref_{max} \end{split}
```

Figure 1: Algorithm IFCSP-SCHEME.

Figures 1 and 2 show the pseudo-code of the general scheme for solving IFCSPs. There are three algorithms: ISCSP-SCHEME, BBE and BB. ISCSP-SCHEME takes as input an IFCSP P and the values for the three parameters: Who, What and When. It returns a partial completion of P that has some necessarily optimal solutions, one of these necessarily optimal solutions, and its preference value. It starts by computing via branch and bound (algorithm BB) an optimal solution of  $P_0$ , say  $s_{max}$ , and its preference  $pref_{max}$ . Next, procedure BBE is called. If BBEsucceeds, it returns a partial completion of P, say Q, one of its necessarily optimal solutions, say  $s_1$ , and its associated preference  $pref_1$ . Otherwise, it returns a solution equal to nil. In the first case the output of IFCSP-SCHEME coincides with that of BBE, otherwise IFCSP-SCHEME returns  $P_0$ , one of its optimal solutions, and its preference.

Procedure BBE takes as input the same values as IFCSP-SCHEME and, in addition, a solution *sol* and a preference

lb representing the current lower bound on the optimal preference value. Function nextVariable, applied to the 1-completion of the IFCSP, returns the next variable to be assigned. The algorithm then assigns a value to this variable. If the Boolean function nextValue returns true (if there is a value in the domain), we select a value for currentVar according to the value of parameter Who.

Function *UpperBound* computes an upper bound on the preference of any completion of the current partial assignment: the minimum over the preferences of the constraints involving only variables that have already been instantiated.

If When = tree, elicitation is handled by procedure Elicit@tree, and takes place only at the end of the search over the 1-completion. The user is not involved in the value assignment steps within the search. At the end of the search, if a solution is found, the user is asked either to reveal all the preferences of the incomplete tuples in the solution (if What=all), or only the worst one among them (if What=worst). If such a preference is better than the best found so far, BBE is called recursively with the new best solution and preference.

If When = branch, BB is performed only once. The user may be asked to choose the next value for the current variable being instantiated. Preference elicitation, which is handled by function Elicit@branch, takes place during search, whenever all variables have been instantiated and the user can be asked either to reveal the preferences of all the incomplete tuples in the assignment (What=all), or the worst preference among those of the incomplete tuples of the assignment (What=worst). In both cases the information gathered is sufficient to test such a preference value against the current lower bound.

If When = node, preferences are elicited every time a new value is assigned to a variable and it is handled by procedure Elicit@node. The tuples to be considered for elicitation are those involving the value which has just been assigned and belonging to constraints between the current variable and already instantiated variables. If What = all, the user is asked to provide the preferences of all the incomplete tuples involving the new assignment. Otherwise if What = worst, the user provides only the preference of the worst tuple.

**Theorem 1.** Given an IFCSP P and a consistent set of values for parameters When, What and Who, Algorithm IFCSP-SCHEME always terminates, and returns an IFCSP  $Q \in PC(P)$ , an assignment  $s \in NOS(Q)$ , and its preference in Q.

If When = tree, then we elicit after each BB run, and it is proven in (Gelain et al. 2007) that IFCSP-SCHEME never elicits preferences involved in solutions which are not possibly optimal. This is a desirable property, since only possibly optimal solutions can become necessarily optimal. However, the experiments will show that solvers satisfying such a desirable property are often out-performed in practice.

## Problem generator and experimental design

We generate IFCSPs via a generator with the following parameters:

• n: number of variables;

```
BBE (P, nInstVar, Who, What, When, sol, lb)
sol' \leftarrow sol, pref' \leftarrow lb
currentVar \leftarrow nextVariable(P_1)
While (nextValue(currentVar, Who))
      If (When = node)
         P, pref \leftarrow Elicit@Node(What, P, currentVar, lb)
      ub \leftarrow UpperBound(P_1, currentVar)
      If (ub > lb)
         If (nInstvar = number \ of \ variables \ in \ P)
           If (When = branch)
             P, pref \leftarrow Elicit@branch(What, P, lb)
           If (pref > lb)
             sol \leftarrow qetSolution(P_1)
             lb \leftarrow pref(P_1, sol)
         else
           BBE(P, nInstVar + 1, Who, What, When, sol, lb)
If (When=tree \ and \ nInstVar = 0)
  If(sol = nil)
   sol \leftarrow sol', pref \leftarrow pref'
    P, pref \leftarrow Elicit@tree(What, P, sol, lb)
   If(pref > pref')
     BBE(P, 0, Who, What, When, sol, pref)
   else BBE(P, 0, Who, What, When, sol', pref')
```

Figure 2: Algorithm BBE.

- m: cardinality of the variable domains;
- d: density, that is, the percentage of binary constraints present in the problem w.r.t. the total number of possible binary constraints that can be defined on n variables;
- t: tightness, that is, the percentage of tuples with preference 0 in each constraint and in each domain w.r.t. the total number of tuples (m² for the constraints, since we have only binary constraints, and m in the domains);
- *i*: incompleteness, that is, the percentage of incomplete tuples (that is, tuples with preference?) in each constraint and in each domain.

Given values for these parameters, we generate IFCSPs as follows. We first generate n variables and then d% of the n(n-1)/2 possible constraints. Then, for every domain and for every constraint, we generate a random preference value in (0,1] for each of the tuples (that are m for the domains, and  $m^2$  for the constraints); we randomly set t% of these preferences to 0; and we randomly set i% of the preferences as incomplete.

Our experiments measure the *percentage of elicited preferences* (over all the missing preferences) as the generation parameters vary. Since some of the algorithm instances require the user to suggest the value for the next variable, we also show the *user's effort* in the various solvers, formally defined as the number of missing preferences the user has to consider to give the required help.

Besides the 16 instances of the scheme described above, we also considered a "baseline" algorithm that elicits preferences of randomly chosen tuples every time branch and bound ends. All algorithms are named by means of the three parameters. For example, algorithm DPI.WORST.BRANCH has parameters Who=dpi,

What=worst, and When=branch. For the baseline algorithm, we use the name DPI.RANDOM.TREE.

For every choice of parameter values, 100 problem instances are generated. The results shown are the average over the 100 instances. Also, when it is not specified otherwise, we set n=10 and m=5.

#### Results

**Incomplete fuzzy CSPs.** The names of all the algorithms and the corresponding line symbols are shown below:

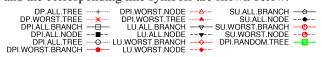


Figure 3 shows the percentage of elicited preferences when we vary respectively the incompleteness, the density, and the tightness. For reasons of space, we show only the results for specific values of the parameters. However, the trends observed here hold in general. It is easy to see that the best algorithms are those that elicit at the branch level. In particular, algorithm SU.WORST.BRANCH elicits a very small percentage of missing preferences (less than 5%), no matter the amount of incompleteness in the problem, and also independently of the density and the tightness. This algorithm outperforms all others, but relies on help from the user. The best algorithm that does not need such help is DPI.WORST.BRANCH. This never elicits more than about 10% of the missing preferences. Notice that the baseline algorithm is always the worst one, and needs nearly all the missing preferences before it finds a necessarily optimal solution. Notice also that the algorithms with What=worst are almost always better than those with What=all, and that When=branch is almost always better than When=node or When=tree.

Figure 4 (a) shows the user's effort as incompleteness varies. As predictable, the effort grows slightly with the incompleteness level, and it is equal to the percentage of elicited preferences only when What=all and Who=dp or dpi. For example, when What=worst, even if Who=dp or dpi, the user has to consider more preferences than those elicited, since to point to the worst preference value the user needs to check all of them (that is, those involved in a partial or complete assignment). DPI.WORST.BRANCH requires the user to look at 60% of the missing preferences at most, even when incompleteness is 100%.

Figure 4 (b) shows the percentage of elicited preferences over all the preferences (white bars) and the user's effort (black bars), as well as the percentage of preferences present at the beginning (grey bars) for DPI.WORST.BRANCH. Even with high levels of incompleteness, this algorithm elicits only a very small fraction of the preferences, while asking the user to consider at most half of the missing preferences.

Figure 4 (c) shows LU.WORST.BRANCH, where the user is involved in the choice of the value for the next variable. Compared to DPI.WORST.BRANCH, this algorithm is better both in terms of elicited preferences and user's effort (while SU.WORST.BRANCH is better only for the elicited preferences). We conjecture that the help the user

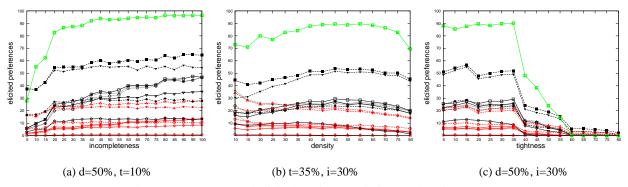


Figure 3: Percentage of elicited preferences in incomplete fuzzy CSPs.

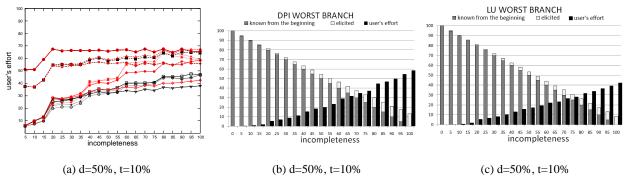


Figure 4: Incomplete fuzzy CSPs: user's efforts and best algorithms.

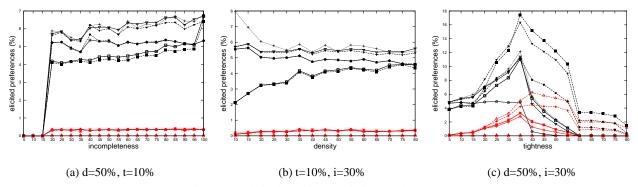


Figure 5: Elicited preferences in incomplete CSPs.

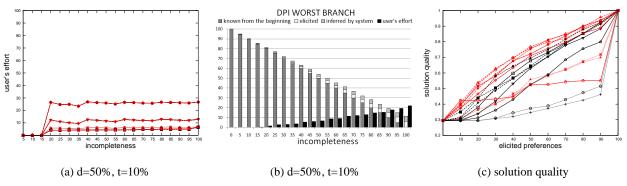


Figure 6: Incomplete CSPs: (a) user's effort and (b) best algorithm. Solution quality for incomplete fuzzy CSPs in (c).

gives in choosing the next value guides the search towards better solutions, thus resulting in an overall decrease of the number of elicited preferences.

Although we are mainly interested in the amount of elicitation, we also computed the time to run the 16 algorithms. The best algorithms in this respect need about 200 ms for problems with 10 variables and 5 elements in the domains, no matter the amount of incompleteness. Most of the algorithms need less than 500 ms.

These algorithms have a useful anytime property, since they can be stopped even before their end obtaining a possibly optimal solution with preference value higher than the solutions considered up to that moment. Figure 6 (c) shows how fast the various algorithms reach optimality. The y axis represents the solution quality during execution, normalized to allow for comparison among different problems. The algorithms that perform best in terms of elicited preferences, such as DPI.WORST.BRANCH, are also those that approach optimality fastest. We can therefore stop such algorithms early and still obtain a solution of good quality in all completions.

**Incomplete hard CSPs.** We also tested these algorithms on hard CSPs. In this case, preferences are only 0 and 1, and necessarily optimal solutions are complete assignments which are feasible in all completions. The problem generator is adapted accordingly. The parameter What now has a specific meaning: What=worst means asking if there is a 0 in the missing preferences. If there is no 0, we can infer that all the missing preferences are 1s.

Figure 5 shows the percentage of elicited preferences for CSPs in terms of amount of incompleteness, density, and tightness. Notice that the scale on the y axis varies to include only the highest values. The best algorithms are those with What=worst, where the inference explained above takes place. It is easy to see a phase transition at about 35% tightness, which is when problems pass from being solvable to having no solutions. However, the percentage of elicited preferences is below 20% for all algorithms even at the peak.

Figure 6 (a) shows the user's effort for the case of CSPs. Overall, the best algorithm is again DPI.WORST.BRANCH, whose performance is shown in Figure 6 (b) in detail.

Incomplete temporal fuzzy CSPs. We also performed some experiments on fuzzy simple temporal problems (Khatib et al. 2007). These problems have constraints of the form  $a \le x - y \le b$  modelling allowed time intervals for durations and distances of events, and fuzzy preferences associated to each element of an interval. We have generated classes of such problems following the approach in (Khatib et al. 2007), adapted to consider incompleteness. While the class of problems generated in (Khatib et al. 2007) is tractable, the presence of incompleteness makes them intractable in general. Figure 7 shows that even in this domain it is possible to find a necessarily optimal solution by asking about 10% of the missing preferences, for example via algorithm DPI.WORST.BRANCH.

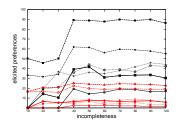


Figure 7: Percentage of elicited preferences in incomplete fuzzy temporal CSPs.

### **Future work**

We plan to consider incomplete weighted constraint problems as well as different heuristics for choosing the next variable during the search. Moreover, we intend to build solvers based on local search or variable elimination methods. Finally, we want to add elicitation costs and to use them also to guide the search, as done in (Wilson, Grimes, and Freuder 2007) for hard CSPs.

### References

Bistarelli, S.; Montanari, U.; and Rossi, F. 1997. Semiring-based constraint solving and optimization. *JACM* 44(2):201–236.

Boutilier, C.; Bacchus, F.; and Brafman, R. I. 2001. UCP-networks: A directed graphical representation of conditional utilities. In *UAI '01*, 56–64. Morgan Kaufmann.

Braziunas, D., and Boutilier, C. 2006. Preference elicitation and generalized additive utility. In *AAAI*. AAAI Press.

Dechter, R., and Dechter, A. 1988. Belief maintenance in dynamic constraint networks. In *AAAI*, 37–42.

Faltings, B., and Macho-Gonzalez, S. 2002. Open constraint satisfaction. In *CP*, volume 2470 of *LNCS*, 356–370. Springer.

Faltings, B., and Macho-Gonzalez, S. 2003. Open constraint optimization. In *CP*, volume 2833 of *LNCS*, 303–317. Springer.

Faltings, B., and Macho-Gonzalez, S. 2005. Open constraint programming. *AI Journal* 161(1-2):181–208.

Gelain, M.; Pini, M. S.; Rossi, F.; and Venable, K. B. 2007. Dealing with incomplete preferences in soft constraint problems. In *Proc. CP'07*, volume 4741 of *LNCS*, 286–300. Springer.

Khatib, L.; Morris, P.; Morris, R.; Rossi, F.; Sperduti, A.; and Venable, K. B. 2007. Solving and learning a tractable class of soft temporal problems: theoretical and experimental results. *AI Communications* 20(3).

Lamma, E.; Mello, P.; Milano, M.; Cucchiara, R.; Gavanelli, M.; and Piccardi, M. 1999. Constraint propagation and value acquisition: Why we should do it interactively. In *IJCAI*, 468–477.

Wilson, N.; Grimes, D.; and Freuder, E. C. 2007. A cost-based model and algorithms for interleaving solving and elicitation of csps. In *Proc. CP'07*, volume 4741 of *LNCS*, 666–680. Springer.