

Position Paper: Explanations for Human-Aware CP  
Eugene C. Freuder  
Insight Centre for Data Analytics  
University College Cork  
Cork, Ireland

Human-Aware AI was the IJCAI-16 Special Theme. As AI becomes more ubiquitous, human-awareness becomes more crucial to ensuring its benefits and avoiding its dangers. (The *PCWorld* report on the IJCAI theme was headlined “How 'human-aware' AI could save us from the robopocalypse”.) A key aspect of human-awareness is the ability of our machines to explain their behavior and their decisions.

Accordingly we focus here on explanation in CP *for users*. Much of the work on explanation in CP has in fact been focused elsewhere, but may prove relevant nonetheless. The most attention has been paid to programs providing explanations for themselves, to enable them to function more efficiently. There has been work on providing explanations to programmers to facilitate development or debugging. There has even been some work on users providing explanations to programs, to facilitate problem acquisition.

Most work on CP explanation has been directed at providing explanation for failure. A successful solution is easily “explained” by observing that all the constraints are satisfied. Nonetheless there may be occasions when we want an explanation for how the solution was obtained: e.g. to indicate why one solution was chosen over another, or to teach users how to find solutions for themselves.

Overall there has not been a great deal of work on explanation in CP. The CP Archive only returns 10 papers from CP conferences with “explanation(s)” in the title (there are none at this year’s CP). So there may well still be some “low hanging fruit”!

A research agenda might include (further) work on explanations for:

- specific classes of CSPs (e.g. distributed CSPs)
- specific application areas (e.g. scheduling problems)
- specific user questions (e.g. “why can’t I have a solution using x”)

There are basic questions about the nature of explanation that merit (further) study, e.g. what makes an explanation “good” or better than an alternative?

As this is a short position paper and not a survey, I will simply cite here some work I have been involved in on user-focused explanation:

1. Inference-based constraint satisfaction supports explanation. M. Sqalli, E. Freuder. *AAAI/IAAI*, Vol. 1, 318-325. 1996
2. A case study in explanation and implication. E. Freuder, C. Likitvivanavong, R. Wallace. *CP2000 Workshop on Analysis and Visualization of Constraint Programs and Solvers*. 2000
3. Deriving explanations and implications for constraint satisfaction problems. E. Freuder, C. Likitvivanavong, R. Wallace. *CP2001*. 2001
4. Explanation and implication for configuration problems. E. Freuder, C. Likitvivanavong, R. Wallace. *IJCAI 2001 Workshop on Configuration*, 31-37. 2001
5. Explanations for whom. R. Wallace, E. Freuder. *CP 2001 Workshop on User-Interaction in Constraint Satisfaction*. 2001
6. Computing explanations and implications in preference-based configurators. E. Freuder, C. Likitvivanavong, M. Moretti, F. Rossi, R. Wallace. *Recent advances in constraints*, 76-92. 2003
7. Generating corrective explanations for interactive constraint satisfaction. B. O’Callaghan, B. O’Sullivan, E. Freuder. *CP05*. 445-459. 2005