Causality, Responsibility, and Blame: A Structural-Model Approach

Joseph Y. Halpern

Computer Science Department
Cornell University
Ithaca, NY 14853, USA
halpern@cs.cornell.edu

This talk will provide an overview of work that I have done with Hana Chockler and Judea Pearl [1,4,5] on defining notions such as *causality*, *explanation*, *responsibility*, and *blame*. I first review the Halpern-Pearl definition of causality—what it means that A is a cause of B—and show how it handles well some standard problems of causality. This definition is based on what are called *structural equations*, which are ways of describing the effects of interventions. The definition (like most in the literature) views causality as an all-or-nothing concept. Either A is a cause of B or it is not. I show how the account can be extended to take into account the degree of responsibility of A for B. For example, if someone wins an election 11–0, each person is less responsible for his victory than if he had won 6–5. Finally, I discuss more recent work [2,3] on combining a theory of normality (or defaults) with the structural equations. A slightly revised definition of causality that uses normality deals well with problems that have been pointed out in the original Halpern-Pearl definition, and helps explain different intuitions that people have regarding causality.

References

1. Chockler, H., Halpern, J.Y.: Responsibility and blame: A structural-model approach. Journal of A.I. Research 20, 93–115 (2004)
2. Halpern, J.Y.: Defaults and normality in causal structures. In: Principles of Knowledge Representation and Reasoning: Proc. Eleventh International Conference (KR 2008), pp. 198–208 (2008)
3. Halpern, J.Y., Hitchcock, C.: Graded causation and defaults (2011) (unpublished manuscript)
4. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach. Part I: Causes. British Journal for Philosophy of Science 56(4), 843–887 (2005)
5. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach. Part II: Explanations. British Journal for Philosophy of Science 56(4), 889–911 (2005)

S. Benferhat and J. Grant (Eds.): SUM 2011, LNAI 6929, p. 1, 2011.