

## ÉLÉMENT DE PORTFOLIO 04



### Publication

#### 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** Action languages based actual causality in ethical decision making contexts

**URL de l'élément :** <https://hal.science/hal-03790699v1>

#### 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Il s'agit d'un article publié dans PRIMA 2022, dont une version journal étendue est en cours de soumission. Il propose une sémantique de causalité réelle basée sur un langage d'action. Cet article est représentatif des récents travaux de thèse de Camilo Sarmiento et pose les bases d'un approfondissement du cadre ACE de raisonnement éthique développé dans l'équipe avec une approche plus formelle et expressive.

#### 3 PRÉSENTATION DE CET ÉLÉMENT

Cet article propose un langage d'action expressif permettant de modéliser des préconditions riches, des actions concurrentes et le déclenchement d'événements exogènes tout en conservant des événements déterministes et instantanés. Il peut être vu un comme un fragment de PDDL+ évitant la complexité des actions duratives. Il permet de déterminer l'évolution du monde étant donné un ensemble d'actions choisies délibérément par des agents et dont l'occurrence peut entraîner une réaction en chaîne d'événements dit exogènes. Cette possibilité d'avoir des réactions en chaîne, allié à la concurrence des actions, permet de représenter finement des situations causales complexes.

Sur la base de ce langage d'action, l'article présente une définition sémantique formelle de causalité réelle basée sur la notion de NESS cause directe (Necessary Element of a Sufficient Set) et discute du bien-fondé de cette approche par rapport aux classiques approches contrefactuelles en montrant que le fait de se baser sur un langage d'action où les effets à court terme fondent la causalité, il est possible de réfuter les principales critiques apportées aux approches causales par régularité. La NESS cause directe définie entre des événements et une formule d'état (formule basé sur des fluents) est ensuite élargie à une notion de NESS cause (non forcément directe) permettant d'aller au-delà des événements déclencheurs directs pour remonter aux actions ayant contribué à la réalisation de la formule. Cette notion est enfin utilisée pour définir une notion de causalité réelle entre deux événements, en se fondant sur les conditions de déclenchement de l'événement causé.

La version étendue de cet article propose de plus une traduction correcte et complète de cette sémantique en ASP, permettant le calcul de toutes les relations causales existantes entre les différents événements d'un scénario donné, permettant ainsi de remplacer le modèle causal initial du cadre de raisonnement éthique ACE développé dans l'équipe (auparavant limité à des théories d'actions exprimées en STRIPS) par ce mécanisme plus fin et expressif.