Robotics: ontology, expressions, propositions, Dennett's physics-design-intention levels, plans, autonomies, obligations,...

Chris Nowak

September 7, 2018

- my interests:

  - reasoning about obligations (ethical reasoning, deontic logics)
  - ethical agents (ethical minds)
  - ethical robots (an ethical mind + its effectors... in a single "body")

- robots in blocks world

  - physical level: wheels, ground, friction,... (below the mind level?)
  - design level: `can-see` and `can-move` blocks,...
  - intention level: `beliefs`, `intentions`, `plans`; `plan-executions`...
    but:
    what is a "plan" and what is a "behaviour"...?
    e.g.:
    `plan-1`: move object `c1` from location `s1` to location `s2`
    `behaviour-1`: execution of `plan-1`
    (`behaviour-1` includes a fragment of the world containing `c1` at location `s1` and a fragment of the world (at a later time) containing `c1` at location `s2`)
    hence,
    plans are in the robot's head, or mind
    behaviours are fragments of the observable world
    (by observing the world, we can guess (deduce) the plans (that have been executed as behaviours))

- are expressions and propositions "abstract" or "physical"...?

  - a cat sat on a mat; the cat sat on the mat; the cat `c1` sat on `m1`...
  - object `c1` is located (at time `t1`) at region `s1` (proposition `p`)
  - the above *proposition* (possibly *true* in the real world) can have "instances" which are fragments of the physical world:
    `instance-1`: the sentence written on the whiteboard
    `instance-2`: the robot's physical representation (of read sentence)

  compare:

  - `honda` is *abstract*, but instances of the concept `honda` are *physical*:
    `instance-1`: the honda car garaged at location garage-1
    `instance-2`: the honda car garaged at location garage-2

- from "intentional level" to "social level"...

- from "intentional level" to "social level"

  - a *single* agent case

    $B\alpha$
    B is a modal *belief* operator (cf. doxastic modal logics)
    α is a proposition
    α could be the proposition: `c1` is located (at time `t1`) at region `s1`

    $I\beta$
    I is a modal *intention* operator (cf. BDI logics)
    β is a proposition
    β could be the prop.: (it is requested) `move c1` from `s1` to `s2`

    in a single agent case B is not subscripted; but if the agent is named, say `b1`, we could/would say: $B_{b_1}\alpha$

  - a *multi* agent case

    B needs to be subscripted; for instance, given agents `b1` and `a1`:
    $B_{b_1}\alpha$ says: agent (or robot) `b1` believes that α
    $B_{a_1}\alpha$ says: agent (or robot) `a1` believes that α
    $B_{b_1}B_{a_1}\alpha$ says: agent `b1` believes that agent `a1` believes that α
    $I_{b_1}B_{a_1}\alpha$ says: the intention of `b1` is that `a1` believes that α

    hence, "social" could be seen as "intentional, in multi-agent worlds". . .

    but note that:
    $I_{b_1}B_{a_1}\alpha$ could be interpreted as *influence*. . .
    $I_{b_1}B_{a_1}O\alpha$ could be interpreted as *control* (O means *obligatory*)

    multi-agent worlds require substantial conceptual analysis. . .
    concepts: *conflict, negotiation, contract, influence, control, aliance, joint goals, collaboration, permissions, obligations, legal / ethical / social constraints, morality, trust,*. . .

    intentional / social capabilities $\rightsquigarrow$ robots with stronger autonomy. . .
    robots that can influence others are more autonomous
    robots that can handle ethical reasoning are more autonomous
    . . .