

Teleology for Joint Action

According to what Csibra and Gergely call ‘the principle of rational action’,

‘an action can be explained by a goal state if, and only if, it is seen as the most justifiable action towards that goal state that is available [to the agent] within the constraints of reality.’¹

This principle can be used to predict an action given knowledge of a goal and the constraints. (It can also be used to infer goals from actions plus constraints, and constraints from actions plus goals (Csibra 2003).) But it doesn’t generally generate predictions where two or more agents are acting together.

To see why not, suppose that Ayesha and Bert are performing actions directed to the goal of moving the table from here to there. How could we use the principle of rational action to predict Ayesha’s action? We might try to treat Bert and his action as falling under ‘the constraints of reality’. But this won’t work if we suppose, further, that Ayesha’s and Bert’s actions are reciprocally coordinated. In that case, which is the ‘most justifiable’ action for Ayesha depends on which is the ‘most justifiable’ action for Bert; and conversely. So we can’t determine which action is the ‘most justifiable’ for Ayesha and separately determine the same for Bert. This is why the principle of rationality as formulated generates no prediction for joint actions which essentially involve reciprocal coordination.

But there is a natural way to generalise the principle of rationality for one or more agents:

One or more actions, $\langle a_1, a_2, \dots a_n \rangle$, can be explained by a goal state if, and only if, they are seen as the most justifiable action towards that goal state that are available to agents $S_1, S_2 \dots S_n$ within the constraints of reality.

This principle can be used to predict actions given knowledge of a distributive goal, the constraints and the fact that the agents are cooperative with respect to the distributive goal. (Knowledge that the agents are cooperative with respect to the goal is necessary to use the principle for predicting their actions: if the agents were competing, the principle would generate incorrect predictions.)

What’s the difference between the original principle of rationality and the generalised version? Where the original concerns the fit between a single action

¹ Csibra & Gergely (1998, p. 255); cf. Csibra et al. (2003). I think this principle needs refining (Butterfill 2012, pp. 6–7), but the refinements don’t matter here. A related but different ‘principle of efficiency’ has been formulated by Southgate et al. (2008, p. 1061): ‘goal attribution requires that agents expend the least possible amount of energy within their motor constraints to achieve a certain end.’ I’m not sure whether differences between this principle and the principle of rationality matter here.

to an outcome, the generalisation concerns the fit between a set of one or more actions and a goal.

Infants use the original principle in predicting individual actions. Will they also use the generalised principle in predicting joint actions?

References

- Butterfill, S. (2012). Interacting mindreaders. *Philosophical Studies*, *forthcoming*, 1–23.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions: Biological Sciences*, *358*(1431), 447–458.
- Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133.
- Csibra, G. & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255–259.
- Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059–1069.