Kraemer's Puzzle and the Theory of Intentional Action

Kyle Blumberg[1] *John Hawthorne*[2]

Arrow: Jane is given the opportunity to push a button which will send a lethal arrow shooting down one of ten specified paths. Jane has no idea which path the arrow will travel down if she pushes the button. But she does know that Bill is standing on path three. Jane hates Bill and wants him to die. So, she pushes the button, the arrow is shot down path three, and Bill is killed.

(1) a. # Jane intentionally shot the arrow down path three.
 b. Jane intentionally killed Bill.

Our informants all agreed that (1a) seems false, but (1b) seems true, even though Jane knows that Bill dies just in case the arrow is shot down path three. We call this *Kraemer's puzzle* in the theory of intentional action, since a similar contrast was first discussed by Kraemer (1978).[3] We provide a solution by formulating a necessary condition on the truth of intentionality reports.

We propose that the key to understanding Kraemer's puzzle is that intentionality reports have a contrastivist aspect: they are *alternative-sensitive.* More precisely, we say that $\mathcal{A}$ is a set of *alternatives* if it is a set of pairwise incompatible propositions. So, if $A, B \in \mathcal{A}$, then $A \cap B = \emptyset$. We maintain that the set of objects that is relevant for the evaluation of an intentionality ascription ⌜S intentionally V⌝ is a set of salient alternatives. We also maintain that ⌜S intentionally V⌝ is true relative to a set of alternatives $\mathcal{A}$ only if S's actions, the outcome V, and the set of alternatives $\mathcal{A}$ are all systematically related. More specifically, we propose the following general constraint: S's basic action must have raised the probability of the V-entailing alternatives in $\mathcal{A}$ substantially more than the $\neg V$-entailing alternatives in $\mathcal{A}$. When this condition is satisfied, we will say that S's basic action *supported* V (relative to a set of alternatives $\mathcal{A}$).

Support condition (rough version): ⌜S intentionally V⌝ is true relative to a set of alternatives $\mathcal{A}$ only if S's basic action raised the probability of the V-entailing alternatives in $\mathcal{A}$ substantially more than the probability of the $\neg V$-entailing alternatives in $\mathcal{A}$.

The support condition features a notion of "raising" the probability of an alternative. We can understand this in terms of a comparative conditional probability calculation. More precisely, given an alternative $B \in \mathcal{A}$, we calculate the difference between the probability of B conditional on S's basic action and the probability of B conditional on S performing the "default action" (both determined at the time of the decision). For simplicity, one can think of the default action as one where the agent does nothing at all. For instance, consider the set of alternatives $\mathcal{A}_{\text{paths}} = \{\textsc{one}, \textsc{two}, \textsc{three}, ..., \textsc{ten}\}$, where ONE is the proposition that the arrow is shot down path one, TWO is the proposition that the arrow is shot down path two, etc. Then in order to determine whether the probability of, e.g. THREE is raised by Jane's basic action of pushing the button, we calculate (i) the probability of THREE conditional on Jane pushing the button, i.e. $\Pr(\textsc{three} \mid \textit{button is pushed})$; and (ii) the probability of THREE conditional on Jane doing nothing (which is equivalent to her not pushing the button), i.e. $\Pr(\textsc{three} \mid \neg\textit{button is pushed})$. In the *Arrow* scenario, $\Pr(\textsc{three} \mid \textit{button is pushed}) = \frac{1}{10}$ and $\Pr(\textsc{three} \mid \neg\textit{button is pushed}) = 0$. So, the amount by which Jane's basic action raised the probability of THREE is $\frac{1}{10} - 0 = \frac{1}{10}$. The support condition asks us to calculate these amounts for each V-entailing alternative, and for each $\neg V$-entailing alternative, and check that the former are greater than the latter. More explicitly:

> **Support condition** (precise version): ⌜S intentionally V⌝ is true relative to a set of alternatives $\mathcal{A}$ only if for all V-entailing $B \in \mathcal{A}$, and for all $\neg V$-entailing $C \in \mathcal{A}$:

[1]Dianoia Institute of Philosophy, Australian Catholic University, Kyle.Blumberg@acu.edu.au

[2]Dianoia Institute of Philosophy, Australian Catholic University; School of Philosophy, University of Southern California, jhawthor@usc.edu

[3]Also see Nadelhoffer (2004) for experimental confirmation of the contrast.

$$\Pr(B \mid S \text{ performs basic action}) - \Pr(B \mid S \text{ does nothing})$$
$$\gg$$
$$\Pr(C \mid S \text{ performs basic action}) - \Pr(C \mid S \text{ does nothing})$$

Let us illustrate with the *Arrow* scenario. Suppose that the set of alternatives relevant for the evaluation of (1a) is $\mathcal{A}_{\text{paths}}$ from above. Then Jane's pushing the button did not support the arrow being shot down path three relative to $\mathcal{A}_{\text{paths}}$. This is because Jane's action raised the probability of *all* of the alternatives in $\mathcal{A}_{\text{paths}}$ equally. As shown above, Jane's basic action raised the probability of THREE by $\frac{1}{10}$. But her basic action raised the probability of every other alternative by $\frac{1}{10}$ as well. Clearly these alternatives entail that the arrow was not shot down path three. So, (1a) can't be true.

By contrast, consider (1b). Suppose that the set of alternatives relevant for this report is $\mathcal{A}_{\text{kill}} = \{\textsc{kill}, \overline{\textsc{kill}}\}$, where KILL is the proposition that Jane kills Bill, and $\overline{\textsc{kill}}$ is the proposition that Jane does not kill Bill. Then Jane's pushing the button does support her killing Bill relative to $\mathcal{A}_{\text{kill}}$. This is because Jane's action raised the probability of KILL: $\Pr(\textsc{kill} \mid \textit{button is pushed}) = \frac{1}{10}$ and $\Pr(\textsc{kill} \mid \neg\textit{button is pushed}) = 0$. On the other hand, Jane's action lowered the probability of $\overline{\textsc{kill}}$: $\Pr(\overline{\textsc{kill}} \mid \textit{button is pushed}) = \frac{9}{10}$ and $\Pr(\overline{\textsc{kill}} \mid \neg\textit{button is pushed}) = 1$. So, assuming that the other conditions on intentional action are satisfied, (1b) is true.

This account also has the potential to explain further contrasts discussed in the literature, e.g. a range of experimental findings from (Malle, 2006). For instance, in one experiment Malle gave subjects the following vignette from (Knobe, 2003), and asked them the questions in (2):

Aunt 1: Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. He raises his rifle, gets her in the sights, and presses the trigger. But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild...Nonetheless, the bullet hits her directly in the heart. She dies instantly.

(2) a. Did Jake intentionally kill his aunt?
    b. Did Jake intentionally hit his aunt's heart?

100% of the respondents answered 'Yes' to (2a). By contrast, only 49% answered 'Yes' to (2b). We can explain this if we suppose that each intentionality report is being evaluated relative to a distinct set of alternatives. For instance, suppose that the set relevant for (2a) is similar to $\mathcal{A}_{\text{kill}}$ and only contains the proposition that Jake kills his aunt and the proposition that Jake does not kill his aunt. Then Jake's basic action, i.e. pulling the trigger, did support him killing his aunt. As for (2b), suppose that this is evaluated relative to the set $\mathcal{A}_{\text{part}} = \{\textsc{heart}, \textsc{lung}, \textsc{kidney}, ...\}$, where HEART is the proposition that Jake hits his aunt's heart, LUNG is the proposition that Jake hits his aunt's lung, etc. Given that Jake has no skill at using the rifle, his basic action does not support hitting his aunt's heart relative to $\mathcal{A}_{\text{part}}$: pulling the trigger raised the probability all of the alternatives in the set equally.

Finally, we explore whether Kraemer effects are exhibited by other constructions. We detect such effects in imperatives, rationale clauses, and control predicates such as 'promise'. For instance, the command 'Shoot the arrow down path three!' sounds much worse than the command 'Kill Bill!'; similarly 'Jane shot the arrow down path three in order to get revenge' is unacceptable, while 'Jane killed Bill in order to get revenge' is felicitous. It has been argued that all of these constructions semantically encode a relation called RESP(ONSIBILITY), where RESP is a two-place relation between an agent S and a proposition p that holds when p follows from some act performed by S with the intention of making p true (Farkas, 1988). In order to explain our observations, we tentatively suggest that the RESP relation itself is alternative-sensitive, and requires the satisfaction of a support condition.

References

Farkas, Donka F. 1988. On Obligatory Control. *Linguistics and Philosophy*, **11**(1), 27–58.

Knobe, Joshua. 2003. Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology*, **16**(2), 309–325.

Kraemer, Eric Russert. 1978. Intentional Action, Chance and Control: [Analysis "Problem" No. 16]. *Analysis*, **38**(3), 116–117.

Malle, Bertram. 2006. Intentionality, Morality, and Their Relationship in Human Judgment. *Journal of Cognition and Culture*, **6**(1-2), 61–86.

Nadelhoffer, T. 2004. The Butler Problem Revisited. *Analysis*, **64**(3), 277–284.