Does causality matter?

Impressions of agency influence judgments of both causal and non-casual sentences

Sehrang Joo [a], Sami R. Yousif [a], Fabienne Martin [c], Frank C. Keil [a], and Joshua Knobe [b]

[a] *Department of Psychology, Yale University*
[b] *Program in Cognitive Science and Department of Philosophy, Yale University*
[c] *ERC LeibnizDream Project, Humboldt-Universität zu Berlin*

Imagine a train platform with a line that people aren't supposed to cross—if they do, incoming trains will automatically stop. Suppose that Tom deliberately steps over the line to stand in front of it, and this ends up causing a train delay. In this case, it seems natural to say:

(1) Tom caused the train delay.

Existing research shows that people's willingness to apply this sentence depends in part on the degree to which Tom is exercising agency. Thus, suppose that, instead of acting intentionally, Tom blacks out and falls over the line. Just as in the first scenario, Tom is now too near the edge of the platform, and this leads to a delay. In this case, however, (1) seems like much less natural way to describe what has happened. Indeed, existing research shows that people's endorsement of sentences like (1) are often affected by whether an agent acted intentionally (see e.g., Kirfel & Lagnado, 2021; Lombrozo, 2010; Rose, 2017; Schwenkler & Sytsma, 2020).

This work typically understands these effects as demonstrating something about *causal* cognition in particular. In other words, existing research has focused especially on judgments about causation and on how impressions of agency might impact those judgments.

Consider, however, the following sentence:

(2) Tom crossed over the line.

In (2), there is no longer any information about causation; the path verb *cross* is typically analyzed as devoid of causative semantics. Yet, strikingly, we find it in the experiments described below that people's evaluations of (2) are affected by intentionality in precisely the same way that their evaluations of (1) are. This result suggests that these effects of intentionality are not about how people reason about causation in particular, but instead show that perceptions of agency impact the way people think about a far broader class of sentences.

This raises a question about what gives rise to the effect of intentionality found in sentences like (1) and (2). One possibility is that these effects are not located in how people reason about the verb in the sentence (i.e., *cause* or *cross*), but instead in how they reason about the subject (i.e., *Tom*). To explore this hypothesis, we can look at cases in which the subject is inanimate:

(3) a. The water caused the train delay.
b. The water crossed over the line.

If these sentences require intentionality in order to be acceptable, then people should also be hesitant to accept (3a-b), since the water is not acting (and cannot act) intentionally. In contrast, if the effect of intentionality has something to do with animate agents in particular, then (3) may be acceptable, since the water is not an animate in the first place.

In our experiments, we find that people endorse (3), to the same extent that they endorse (1) and (2) when Tom acts intentionally. These results suggest that intentionality affects the evaluation only of sentences that are about animate agents (and does so whether or not those sentences involve explicit causation).

Experiment 1

Four hundred adult participants were shown one of four short vignettes about a person, Tom, acting with full agency or with a very low degree of agency. For example, in one vignette, participants were told that Tom is waiting for a train and that there is a yellow line on the platform that people aren't supposed to cross. In the full agency condition, Tom deliberately crosses over the line, causing an adverse outcome. In the reduced agency condition, Tom passes out and falls over the line, causing the same outcome. Participants were then asked to evaluate *either* a causal statement (e.g., "Tom caused the train delay.") *or* a statement with one of the four non-causative verbs *hit, touch, enter* and *cross* (e.g., "Tom crossed the line.") on the basis of whether this sentence was a "natural/valid way of describing the event."

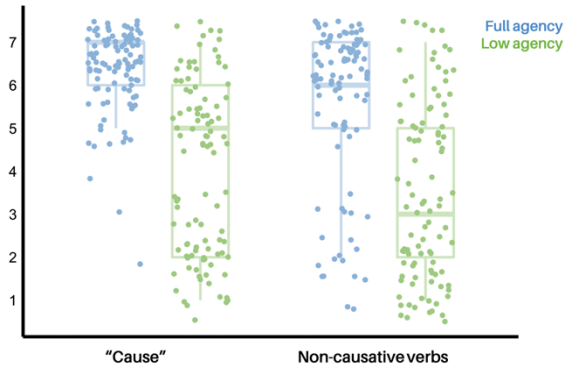


Figure 1. Results from Experiment 1.

Results are displayed in Figure 1. We found no significant interaction between degree of agency and statement type. There was, however, a significant effect of degree of agency within each statement type (*p*s<.001). This means that whether or not Tom acted with full agency affected participants' evaluations of both causal and non-causal statements.

Experiment 2

Six hundred adult participants were again shown one of four short vignettes. Now, however, participants were split into three agency conditions: (1) Tom acting with a very high degree of agency (e.g., Tom, in full control of his actions, deliberately stepping over the line); (2) Tom acting with very low agency (e.g., Tom blacking out and falling over the line); and (3) an inanimate object acting with little to no agency (e.g., a heavy rainstorm floods the train platform, and the weight of the water over the line triggers the same outcome). Participants were again asked to evaluate *either* a causal statement (e.g., "Tom caused the train delay" or "The water caused the train delay") *or* a statement with a non-causative verb (e.g., "Tom crossed over the line" or "The water crossed over the line") on the basis of whether this sentence was a "natural/valid way of describing the event."

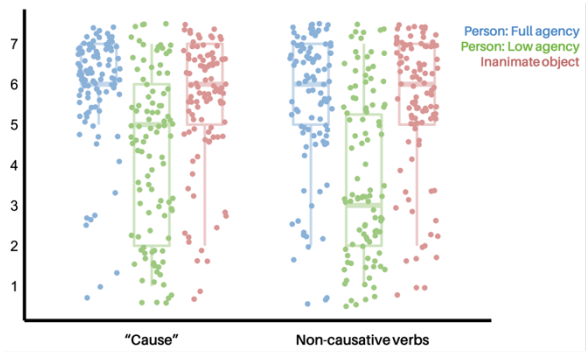


Figure 2. Results from Experiment 2.

Results are displayed in Figure 2. We again found no significant interaction between *degree of agency* and *statement type*—replicating the effect of degree of agency across sentences with both causative and non-causative verbs. Furthermore, degree of agency affected participants evaluations of sentences about Tom, such that sentences describing Tom's actions were rated as more natural/valid when Tom acted intentionally than when he did not (*p*<.001)—but did *not* affect their evaluation of sentences about inanimate objects; participants thought a sentence like "The water crossed over the line" was an acceptable description of the scenario (even though the water obviously had a very low or null degree of agency; *p*=.30).

Conclusion

The effect of intentionality on people's evaluations of sentences like (1) are well-documented. We find, however, that these effects do not arise from something about causal cognition in particular. Instead, they may result from some more general role that agency plays in language. Thus to best understand how people are reasoning about intentional action in these cases, future research should focus not on developing theories that are specific to causal cognition in particular—but instead on developing theories designed to capture more general effects involving the role of agency in language.

DETACHED AGENCY: HOW TO GO THROUGH THE MOTIONS

Jules Salomone-Sehr
Centre de Recherche en Éthique – McGill University
jules.salomone@mcgill.ca

Suppose Φ is a feasible course of behavior. How might you Φ at will? Here is an obvious answer: you form the intention to Φ, and then enact that intention. *Ordinary intentional agency*, however, is not the only mechanism through which you might Φ at will. In this paper, I highlight a more circuitous mechanism, one that dispenses with the intention to Φ, hence that *enables to* Φ *at will without intending to* Φ. I call this mechanism *detached agency*. In highlighting this mechanism, I aim to advance our understanding of agency under non-ideal circumstances.

First, I motivate the need for an alternative to ordinary intentional agency by drawing attention to cases where Φ-ing is something we'd rather not do, but have no choice but to do. In such cases, it would be comforting news if you could get yourself to Φ without having to intend to Φ. If Φ is morally repugnant, for instance, managing to Φ without intending to Φ would keep your intentional psychology safe from involvement in wrongdoing—hence, would arguably reduce your responsibility for Φ-ing. There might even be cases where forming the intention to Φ is simply beyond you and where, therefore, ordinary intentional agency won't enable you to Φ. Imagine a gay man in a heteronormative society. He might have no choice but to marry a woman, and yet be psychologically incapable to form that intention. Here too, an alternative to ordinary intentional agency would be welcome, for ordinary intentional agency won't help that man comply with oppressive social norms (and avoid sanctions for non-compliance).

Second, I explain what detached agency involves as follows:

1/ You map out the steps $\varphi_1, \varphi_2 \ldots, \varphi_n$ which, if completed, would very probably suffice for successful Φ-ing.

2/ You form the intention to enact the plan that comprises steps φ_1 through φ_n, while nonetheless not committing yourself to Φ-ing. So, although committed to a plan likely to result in Φ-ing, you would not care if that plan failed to result in Φ-ing due to changing circumstances.

3/ You enact your plan.

As this suggests, detached agency exploits the gap between intending to Φ and intending to enact a plan designed to Φ. Though you intend the enactment of your plan, you have no investment one way or another in its success: you merely go through the motions. And yet, if your plan is well designed and if the world cooperates, enacting that plan (something you can do at will) enables you to Φ.

Third, I defuse two worries. Detached agency, I argue, need involve neither self-deception, nor irrationality.

There would be self-deception if deliberating about how to Φ, or intending to φ_1 through φ_n implied, as a matter of psychological necessity, forming the intention to Φ. For then detached agency would be no alternative to ordinary intentional agency. However, I argue that you can deliberate about how to Φ while remaining uncommitted to Φ-ing, as when you imagine what a tentative course of action would involve. I also argue that intending to φ_1 through φ_n does not amount to intending to Φ. As intention is normally conceived (e.g., Bratman 1984; 1987), intending to Φ disposes you to check that your plan is on track to Φ, and to adjust your plan should it go off track. Similarly, then, intending to φ_1 through φ_n implies a tracking disposition to $\varphi_1 \ldots$, and φ_n. Now, due to changing circumstances, φ_1-ing…, and φ_n-ing might perhaps no longer suffice for Φ-ing. However, if you merely intend to φ_1 through φ_n, you won't be disposed to adjust your conduct to remain on track to Φ (Kutz 2000: 100-101). In short: the intention to Φ and the intention to enact a plan designed to Φ come with different tracking dispositions, a fact that makes detached agency possible.

There would be irrationality if detached agency, in addition to involving the intention to enact a plan designed to *Φ*, also involved the intention to *not Φ*. For this violates the requirement that intentions be means-end coherent (Bratman 1987; Broome 2013): enacting a plan designed to *Φ* which you know to be so designed is a terrible means towards satisfying your intention to *not Φ*. Detached agency, however, does not involve intending to *not Φ*, but merely *not* intending to *Φ*. There would also be irrationality if rationality required to intend to *Φ* when you already intend to enact a plan designed to *Φ*. But that is not the case. Imagine that you are following the steps of a mayonnaise recipe you created just to see what will happen, perhaps because you are a professional cook and you are interested in assessing how that recipe compares to others. In that scenario, you need not be committed to making mayonnaise: in fact, if your following of the steps of that recipe failed to yield mayonnaise, you would not adjust your conduct—but simply note that this recipe is no good. In this scenario, you intend to enact a plan that you have designed to *Φ*, while not intending to *Φ*, but clearly your conduct cannot be suspected of any practical irrationality. It is therefore false that rationality requires to intend to *Φ* when you intend to enact a plan that you have designed to *Φ*. Lastly, there would be irrationality if detached agency conflicted with the principle of *Intention Agglomeration* according to which (roughly) it is rationally required, when intending to *Ψ* and intending to *X*, to also intend to *Ψ* and *X* (Yaffe 2004; Velleman 2007; Bratman 2009; Goldstein 2016). True, a detached agent intends to enact each step of the plan φ_1…through φ_n; but this only creates a rational pressure to intend to enact the conjunction of φ_1…, and φ_n, and not to intend to *Φ*. For as I argued earlier, the intention to *Φ* is not identical to the intention to enact a plan designed to *Φ*. Detached agency, therefore, does not fly in the face of *Intention Agglomeration*.

Lastly, I show how the concept of detached agency might be put to philosophical work. Focusing on oppression theory, I suggest that detached agency might clarify the debate between those who think that oppressed agents can be fully agentic (Khader 2018; 2020), and those who claim, instead, that oppression jeopardizes the agency of the oppressed (Oshana 2006; Stoljar 2014). Oppression, I contend, is a context where detached agency is made desirable, for it enables you to do whatever is required to comply with oppressive norms while not intending to so comply. Oppression, in other words, is a context where the oppressed have reason to deliberate ordinary intentional agency away and, instead, to engage in detached agency.

Bratman, Michael. 1984. “Two Faces of Intention.” *The Philosophical Review* 93 (3): 375–405.

———. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

———. 2009. “Intention, Belief, Practical, Theoretical.” In *Spheres of Reason: New Essays in the Philosophy of Normativity*, edited by Simon Robertson. Oxford University Press.

Broome, John. 2013. *Rationality Through Reasoning*. John Wiley & Sons.

Goldstein, Simon David. 2016. “A Preface Paradox for Intention.” *Philosophers’ Imprint* 16.

Khader, Serene J. 2018. *Decolonizing Universalism: A Transnational Feminist Ethic*. Oxford University Press.

———. 2020. “The Feminist Case Against Relational Autonomy.” *Journal of Moral Philosophy* 17 (5): 499–526.

Kutz, Christopher. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press.

Markovits, Julia. 2011. “Why Be an Internalist About Reasons?” *Oxford Studies in Metaethics* 6: 255.

Muñoz, Daniel. 2018. “Thinking, Acting, Considering.” *Australasian Journal of Philosophy* 96 (2): 255–70.

Oshana, Marina. 2016. *Personal Autonomy in Society*. Routledge.

Stoljar, Natalie. 2014. “Autonomy and Adaptive Preference Formation.” In *Autonomy, Oppression, and Gender*, edited by Andrea Veltman and Mark Piper. Oxford University Press.

Velleman, J. David. 2007. “What Good Is a Will?” In *Action in Context*, edited by Anton Leist and Holger Baumann. de Gruyter/Mouton.

Yaffe, Gideon. 2004. “Trying, Intending, and Attempted Crimes.” *Philosophical Topics* 32 (1/2): 505–32.

French double Agents: a force-theoretic account of *de* and *par*

Camil Staps (Leiden University / Radboud University Nijmegen) and Johan Rooryck (Leiden University)

In French passives, there is a contrast between Agents that are introduced by the preposition *par* 'by, via' and those that are introduced by *de* 'from'. In the sentences in (1), *par* is the only option, while the sentences in (2) allow for an alternation between *par* and *de*, and those in (3) allow only *de*.

(1) a. *Le chien est lavé par/ *de Marie.* (Straub 1974: 584)
'The dog is washed by Mary.'
b. *La fenêtre a été brisée par/ *d'un rocher.* (Straub 1974: 584)
'The window has been broken by a rock.'

(2) a. *Le professeur était respecté par/ de ses élèves.* (Straub 1974: 585)
'The professor was respected by his students.'
b. *Le garçon est adoré par le/ du grand-père.* (Clédat 1900: 223)
'The boy is loved by his grandfather.'

(3) a. *Le mois de février est précédé du/ *par le mois de janvier.* (Straub 1974: 591)
'The month of February is preceded by the month of January.'
b. *Le dernier chapitre est suivi d'/ *par une table des matières.*
'The last chapter is followed by a table of contents.'

It is generally agreed that passive Agents in French are by default introduced by *par* 'by, via', while the conditions on the use of Agentive *de* 'from' are more intricate and elusive. This distinction has been analysed in terms of style (*de* being more formal than *par,* e.g. Gougenheim 1938: 307) and Aktionsart (*de* only being used with stative verbs, e.g. Zumthor & von Wartburg 1947: 297). However, careful description shows that these analyses cannot account for all cases (Gaatone 1998). Straub (1974) provides the most complete descriptive generalisations so far:

(4) a. The Agent of a verb that brings about a change is always marked by *par* (cf. (1)).
b. Verbs denoting states with animate Agents can be marked by both *de* and *par* (cf. (2)).
c. Verbs denoting states with inanimate Agents always take *de*. (cf. (3))

Generalization (4c) is too restrictive: many examples can be found online of *précédé/ suivi par* 'preceded/ followed by' in the relevant context, contradicting the judgments in (3). Generalization (4b) is not restrictive enough: *de* is not always allowed with verbs that do not involve change of state. This is shown by the contrasts in (5-6):

(5) a. *Les étudiantes sont accompagnées par/de leurs familles.* (after Gaatone 1998: 200)
'The students are accompanied by their families.'
b. *Le détenu est accompagné par le/*du policier.*
'The prisoner is accompanied by the policeman.'

(6) a. *La rock star est toujours suivie d'/ par une foule d'admirateurs.* (Straub 1974a: 25)
'The rock star is always followed by a crowd of admirers.'
b. *Le criminel a été suivi par le / *du détective.* (after Gaatone 1998: 203)
'The criminal has been followed by the detective.'

We propose a formalization of the semantics of *de* and *par* that both simplifies and derives the generalizations in (4), while also capturing the contrast in (5-6). The intuition is as follows. When a prisoner is accompanied by a policeman, that policeman is pragmatically interpreted as exerting force on the prisoner, while students accompanied by their families are not viewed as being subject to a similar force. We therefore propose that in (5b-6b), *de* is excluded because the Agent is viewed as exerting a tangible influence on the prisoner or criminal, whereas the students and the rock star in (5a-6a) remain entirely unaffected by the Agent. This distinction can be captured in terms of Talmy's (1988, a.o.) force dynamics, formalized in the framework of Copley & Harley (2015) as in (7), and less formally in (8):

(7) a. $[\![\text{de}]\!] = \lambda s.\lambda e.\lambda f.\ \text{Source}(f, e) \wedge f(s) = s$

b. [[par]] = λs.λe.λf. Source(f, e)

(8) a. *De* selects DPs naming the source of a force that does not yield a new situation.
b. *Par* selects DPs naming the source of a force that may or may not yield a new situation.

This formulation derives a number of properties of *de* and *par*. First of all, the semantics of *par* is less specific than that of *de*. This derives the observation that *par* is the default option for expressing passive Agents in French. Secondly, *de* carries a specification that it marks a force that does not yield a new situation: this derives the observation that *de* can only be used to refer to Agents that do not influence the Patient. The analysis also accounts for aspectual differences. In (9), the aspectual interpretation of the event varies with the preposition: with *de*, the event describes a continuous state; with *par*, it is inchoative:

(9) *Les élèves sont émerveillés de/par ses découvertes.* (Straub 1974: 590)
'The students are amazed by his/her discoveries.'

This aspectual effect derives from our analysis. The use of *de* requires an interpretation in which the students are in a state of amazement at the discoveries without being changed by them, while *par* demands an interpretation in which the students are being brought into a state of amazement by a force inherent in the discoveries. Finally, the literature suggests that the analysis extends to the counterparts of *de* and *par* in Spanish (Suñer 1981) and Portuguese (Moody 1972).

Regarding the division of labor between *par* and *de*, we argue that when *par* became the default marker for Agent in passives, *de* was reinterpreted as a marker of Agents whose force does not exert an influence over the Patient. The preposition thus came to express 'non-influential involvement' in a situation. This development can be related to the spatial origins of the prepositions: *de* 'from' marks a more remote location than *par* 'by, via'. We suggest that the greater spatial distance expressed by *de* is reinterpreted force-dynamically as a smaller influence.

Finally, we discuss the theoretical implications of our analysis for the formal force-theoretic framework developed by Copley & Harley (2015). Note that we have to distinguish situations with a non-influential force (as with *de*) from situations in which there is no force at all (e.g., with predicates like *resemble*, *be green*). This allows us to account for predicates that require energy input from some Agent but do not affect the Patient, as we see in examples like (5–6). A similar distinction is needed to account for verbs of maintaining (e.g. *keep*, *stay*), which presuppose a force, as opposed to simple *be* (Copley & Harley 2015: 146–150). Our work thus provides additional evidence that the force-theoretic framework needs to be able to distinguish three cases: (a) the lack of a force; (b) the presence of a force that does not effect a change (i.e., does not yield a new situation); (c) the presence of a force that does effect a change (i.e., yields a new situation).

Summarizing, the contributions of our paper are threefold:

1. We provide a simpler yet more accurate account of the prepositions *de* and *par* in French passive sentences, and their cognates in Spanish and Portuguese.
2. We suggest a spatial origin for the development of *de* into a marker of non-influence.
3. We provide evidence for a distinction between the absence of a force and the presence of a force that does not effect a change in the force-theoretic framework of Copley & Harley (2015).

References: Clédat, Léon. 1900. *De* et *par* après les verbes passifs. *Revue de philologie française et de littérature* 14:218–233. | Copley, Bridget, & Heidi Harley. 2015. A Force-Theoretic framework for event structure. *Linguistics & Philosophy* 38 (2): 103–158. | Gaatone, David. 1998. *Le passif en français*. Paris: Duculot. | Gougenheim, Georges. 1938. *Système grammatical de la langue française.* Paris: d'Artrey. | Moody, Raymond. 1972. Portuguese prepositions: some semantic categories. *Luso-Brazilian Review* 9 (1): 36–71. | Straub, Sylvia A. 1974. The passive prepositions *de* et *par*. *The French Review* 47 (3): 583–593. | Suñer, Margarita. 1981. Por vs. De: agential prepositions? *Hispania* 64 (2): 278–283. | Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 12:49–100. | Zumthor, Paul & Walther von Wartburg. 1947. *Précis de syntaxe du français contemporain*. Bern: Francke.

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 $\mathcal{A}_{\text{paths}} = \{\text{ONE}, \text{TWO}, \text{THREE}, ..., \text{TEN}\}$, 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. $\Pr(\text{THREE} \mid \textit{button is pushed})$; and (ii) the probability of THREE conditional on Jane doing nothing (which is equivalent to her not pushing the button), i.e. $\Pr(\text{THREE} \mid \neg\textit{button is pushed})$. In the *Arrow* scenario, $\Pr(\text{THREE} \mid \textit{button is pushed}) = \frac{1}{10}$ and $\Pr(\text{THREE} \mid \neg\textit{button is pushed}) = 0$. 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 $\mathcal{A}_{\text{kill}} = \{\text{KILL}, \overline{\text{KILL}}\}$, where KILL is the proposition that Jane kills Bill, and $\overline{\text{KILL}}$ 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: $\Pr(\text{KILL} \mid \textit{button is pushed}) = \frac{1}{10}$ and $\Pr(\text{KILL} \mid \neg\textit{button is pushed}) = 0$. On the other hand, Jane's action lowered the probability of $\overline{\text{KILL}}$: $\Pr(\overline{\text{KILL}} \mid \textit{button is pushed}) = \frac{9}{10}$ and $\Pr(\overline{\text{KILL}} \mid \neg\textit{button is pushed}) = 1$. 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 $\mathcal{A}_{\text{part}} = \{\text{HEART}, \text{LUNG}, \text{KIDNEY}, ...\}$, 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.

Outcome effects, moral luck and the hindsight bias

Markus Kneer (University of Zurich)
Izabela Skoczeń (Jagiellonian University)

John and Sally drive to work. They are well-rested, alert and stick to the speed limit. A child jumps in front of John's car and dies, Sally arrives at work without incident. Who is more to blame? In between-subjects designs, a pronounced outcome effect tends to arise: John is judged morally and legally more culpable than Sally (henceforth the Outcome Effect). This might strike us as unjust, if we hold, with Kant (1978), that agents are morally responsible only for features of their actions over which they have control (the Control Principle).

Philosophers assume that a difference in moral judgment arises even within-subjects, i.e. when people directly compare John's and Sally's case (the Difference Intuition). This would give rise to the Problem of Resultant Moral Luck: We must square the consequentialist Difference Intuition with the Kantian Control Principle, but the two are fundamentally inconsistent. However, Folk Morality disagrees: When presented with John and Sally's cases side by side, the vast majority of participants evaluate the two agents identically. Western Criminal Law, with its deep distaste for strict liability, sides with the Folk in this regard. So there might not be a complex philosophical problem (the within-subjects Difference Intuition is simply an oddity of philosophers hunting for a paradox). However, in everyday life, we are not confronted with two neat cases side-by-side. Usually, we assess situations where a concrete harm has occurred and here outcome is likely to have distorting effect on our judgment, violating the Control Principle to which both the Law and the Folk are committed.

How can we alleviate the outcome effect? There is evidence in favour of a probabilistic account of moral luck-type phenomena (Kamin & Rachlinski, 1995; Kneer & Machery, 2019). On this account, the post-hoc probability of harming a child is perceived higher for John than for Sally. It thus seems more appropriate to judge that John incurred a substantial risk than that Sally did, which, in turn would mean he was more reckless or negligent than Sally. If this account is on the right track, then a perceived difference in probability and risk drives an asymmetry of risk-related inculpating mental states and hence moral (and legal) evaluation. The whole series of inferences from descriptive features to normative evaluation is innocuous, except for the first step, which is affected by the hindsight bias: in John's case, people tend to exaggerate the degree to which a harmful outcome could, or should, have been anticipated (Fischhoff, 1975; 1980). To address the distorting effect of outcome on culpability judgments, this suggests, we must find ways to alleviate the hindsight bias.

We first explore whether the probabilistic account of the effect of outcome on culpability replicates. Our experiments (total N = **2043**) are the first to control explicitly for the distinction between objective probability (probability from the perspective of the universe) and subjective probability (as perceived from the agent's context). Having replicated the outcome effect on probability, mens rea and moral judgment, we show that it must be considered a bias. The effect of outcome is much more pronounced in between-subjects designs than in within-subjects designs. Next, we turn to debiasing strategies: first, probability anchoring. We test whether giving participants the possibility to evaluate the likelihood of a harmful outcome before the consequences are revealed has an impact on their probability assessments ex post. Next, counterfactual priming: we investigate whether entertaining alternative outcomes reduces the outcome effect on probability, mens rea and moral judgments. Finally, probability stabilizing, in which an expert provides the

actual ex ante probability of a harmful outcome from the point of view of a scientifically informed perspective. Probability anchoring and counterfactual priming attempt to prevent inappropriate inferences from outcome information to probability ex post in indirect fashion. By contrast, probability stabilizing makes short shrift of the problem by directly stipulating the probability ex post so as to prevent inadequate downstream consequences on mens rea and culpability assessment. Consistent with previous research, the effects of outcome on probability post hoc and downstream variables such as mens rea and culpability are persistent and robust across experiments with different scenarios. These effects are the results of a cognitive bias (though not for punishment judgments). Neither strategy fully eradicates inappropriate inferences from outcome to probability and distorted downstream effects on mens rea and culpability judgments thus remain. What works best is probability stabilizing, which is indeed a means courts all too frequently do not resort to.

Selected References:

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Kamin, K. A., & Rachlinski, J. J. (1995). Ex post ≠ ex ante: Determining liability in hindsight. *Law and Human Behavior*, 19(1), 89–104. https://doi.org/10.1007/BF01499075

Kant, I. (2009). *Critique of pure reason* (15th reprint). Cambridge University Press.

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, 182, 331–348. https://doi.org/10.1016/j.cognition.2018.09.003

Aiming at culmination: causal models, event types, and the imperfective paradox

Prerna Nadathur
Universität Konstanz

Elitzur Bar-Asher-Siegal
Hebrew University of Jerusalem

Progressives of telic predicates famously give rise to the **imperfective paradox** (IP; Dowty 1979), in the clash between a standard *culmination assumption* (that uninflected telic predicates exclusively denote culminated eventualities) and the observation that telic progressives are acceptable in culmination-precluding contexts.

(1) Mahler was writing a tenth symphony (when he died). $\not\rightarrow$ *He completed the symphony.*

Given the culmination assumption, prominent accounts of the IP propose intensionalizing the progressive operator, PROG, so that it instantiates qualifying (culminated) P eventualities in modal alternatives to the evaluation world (Dowty, Asher 1992, Bonomi 1997, a.o): thus, an evaluation world eventuality satisfies PROG(P) iff it continues to culmination across the relevant alternatives.

We propose an alternative approach, on which the intensionality relevant for IP effects is not introduced by PROG, but instead inheres in the mereological structure of telic predicates. On our analysis, the truth of a telic progressive depends on a correspondence between reference time facts and an **event type** $[\![P]\!]$, which is structured as a **causal model** for P's culmination condition (C_P). A reference-time token *aims at culmination* iff it represents a plausible 'cross-section' of a causal pathway for C_P as defined by $[\![P]\!]$. This approach delivers improved judgements for challenging IP data, in particular offering an account of contexts in which agents' intentions appear to supersede a realistic assessment of the (im)possibility of culmination.

Culmination, expectation, intention. Culmination alternatives for intensional PROG are usually identified with worlds containing *normal* (or *inertial;* Dowty) developments of a reference-time situation, thus predicting telic progressives to be false whenever C_P is unexpected from the salient *perspective* (cf. Asher). The predictions are reasonable for contexts like (2), ruling out the 'objective' progressives (2a), but permitting (2b) (as assessed from the child's unrealistic perspective).

(2) *Context:* Meena's five-year old daughter Maya wrongly believes that the earth is made entirely of sand and soil. She is digging a hole with the intention of tunnelling through to the other side.
(a) *Meena:* #/?Maya is digging a hole to China. (b) *Maya:* I am digging a hole to China.

However, intensional PROG approach incorrectly predicts the falsity of *out of reach* (OOR) progressives, where an agent's intentions are at odds with their (realistic) assessment of the reference-time accessibility of culmination. (3a-b) are judged to be both acceptable and true, even though his inevitable pre-finish collapse is explicitly included in Benny's perspective as well as that of an objective (but knowledgeable) observer (cf. Szabó 2008).

(3) *OOR context:* Benny began an ultramarathon for which he (knowingly) undertrained; it was certain before the start that he had insufficient stamina to complete the run.
(a) *Objective:* Benny was running an ultramarathon (when he collapsed from exhaustion).
(b) *Benny:* I was running an ultramarathon (when I collapsed).

(Informal) desiderata. OOR data show that the truth of a telic progressive cannot be based solely on a 'local' expectation of culmination, but must take a more general view of the relationship between reference-time facts and the goal at which they aim. What unites (2b)-(3), and differentiates them from (2a) is a *world-historical* possibility of culmination, assessed from the speaker's perspective. In each acceptable example, the speaker associates predicate P with at least one realizable culmination procedure; the progressive assertion reports the belief that reference-time facts correspond to steps along such a completion pathway. On this view, licensed telic progressives are true as long as reference-time facts are compatible with *what would need to be happening* for culmination to take place. This comes apart from the intensional PROG approach in cases like (3), where culmination is contextually but not categorically precluded. Informally speaking, knowledge of an agent's intentions provides evidence of adherence to a culmination procedure: thus, (3a-b) are true because the speaker believes Benny to be (intentionally) doing *what one does* in order to run an ultramarathon.

Causal models for telic event types. The intuition underlying our approach is that telic progressives (agentive or otherwise) report a match between reference-time facts and a culmination procedure for predicate P. We formalize this intuition in terms of the relationship between a reference-time **token** and a (structural equation) **causal model** for the P **event type**. Such models establish generalizations about causal relationships between a finite set Σ of propositional variables by means of a directed acyclic graph with vertices in Σ, accompanied by a set of equations indicating how the value of a particular variable is determined by the values of its ancestors in the graph (Pearl 2000, Schulz 2011).

An event type model M_P for telic predicate P relates conditions (facts, properties of individuals or objects) which are causally relevant for $P's$ culmination condition (dependent variable C_P). Maximal (culminated) P-eventualities correspond to *complete causal pathways* for C_P: that is, sets S of condition-valuation pairs (whose causal interrelationships are provided by M_P) that are jointly sufficient for the truth of C_P ($\text{SUFF}^{M_P}(S, C_P)$). M_P induces a type-level mereological structure where $[\![P]\!]$ contains (non-)culminated eventualities; $e_1, e_2 \in [\![P]\!]$ are comparable if and only if they are partial realizations of the same causal pathway S for C_P.

Truth and felicity of telic progressives. Given a model M_P for telic predicate P with culmination condition C_P, $\text{PROG}(P)$ is true just in case the reference-time situation s is a possible 'cross-section' of a non-culminated P-eventuality: i.e., iff (a) s realizes some part (i.e., some condition Q) of a causal pathway for C_P, (b) does not realize a complete pathway for C_P, and (c) does not realize a sufficient set for non-culmination ($\neg C_P$).

(4) $\text{PROG}(P, t) = 1$ iff $\exists s[\tau(s) \circ t \wedge [\exists Q \exists S : Q \in S \wedge \text{SUFF}^{M_P}(S, C_P) \wedge Q(s)]$ (a)
$\wedge [(\forall S' : \text{SUFF}^{M_P}(S', C_P)[\exists Q' \in S' : Q'(s) \rightarrow \exists Q'' \in S' : \neg Q''(s)]]$ (b)
$\wedge [\forall \Omega : \text{SUFF}^{M_P}(\Omega, \neg C_P)[\exists \omega \in \Omega : \neg \omega(s)]]]$ (c)

Given an epistemic state which supports event type M_P, telic progressives are true of situations which *aim at culmination* (insofar as they have the possibility of continuing to develop along a causal pathway for C_P; cf. Landman 1992). Telic progressives are therefore true not in virtue of the actual consequences of the reference-time situation, but instead in view of type-level causal relationships between reference-time facts and C_P.

This approach captures the empirical judgements in (2)-(3). (2a) is infelicitous (not false) because the speaker's realistic perspective does not admit causal models for physically impossible tasks. By contrast, (2b)-(3) are felicitous and true (from the speaker's perspective) because reference-time facts are compatible with partial P-eventualities in the causal structure of $[\![P]\!]$. By severing the truth of telic progressives from the local accessibility of culmination, the causal approach also allows us to account for the role of agents' intentions. Within models for agentive telic predicates, intentions hold a special status as *globally necessary* conditions for C_P: intent belongs to all sufficient sets for culmination, and insofar as they must be *sustained* through the development of a P-eventuality (cf. Varasdi 2014), their negations are singleton sufficient sets for $\neg C_P$. (4c) thus predicts the falsity of agentive progressives in any context where goal-directed intention fails.

Insofar as sustaining conditions (including intentions and non-agentive analogues like momentum or velocity) provide evidence that licenses comparison between a reference-time situation and event type P, we further suggest that these conditions operate as minimal preconditions for membership in $[\![P]\!]$: telic progressives are thus felicitously used only when (a) the event type model is licensed, and (b) any sustaining conditions in M_P are set in the culmination-conducive way ($\forall Q : \text{SUFF}^{M_P}(\{\neg Q\}, \neg C_P), Q(s)$). Example (5) supports a presuppositional role for intention, showing that Benny's nonspecific intention is enough to make both claims infelicitous, even where his actions otherwise adhere to established procedures for completing either distance.

(5) *Nonspecific intention:* Benny began running in a marathon (42km). Knowing that he had undertrained, he planned to decide at 15km whether to stop there or continue to 21km. He collapsed at 10km.
(a) *Benny:* ?I was running a 15K. (b) *Benny:* ?I was running a half-marathon.

Outlook. It has long been clear that intensional approaches to the IP must be supplemented by a mereological theory which permits the comparison of (non-)culminated telic eventualities (Bach 1986, Landman, Bonomi). We here propose to capture IP effects via a mereological structure which is inherently intensional in that it unifies (non-)maximal P-eventualities in terms of a shared relationship to culmination condition C_P. Causal event type models not only provide a means of measuring the development of telic eventualities in the absence of concrete correlates (such as incremental themes; cf. Parsons 1990), but, through the special status awarded to sustaining conditions within the model, also establish formal criteria for minimal P-eventualities, thus accounting for judgements like (3)-(4), which resist analysis on received intensional PROG approaches. Insofar as the type of data which provides evidence for intention is distinct from that for non-agentive sustaining conditions, we anticipate that the causal approach will shed light on independently-observed agentivity contrasts in the derivation of culmination entailments (see, e.g., Martin & Schäfer 2012 on *defeasible causatives*).

The 'normality' intuition underlying intensional PROG approaches to the IP—i.e., that culmination should represent a 'normal' outcome of P-in-progress—is recognized here as a type-level intuition: $\text{PROG}(P)$ is true of a reference-time situation s because s corresponds to a *normative* path for culmination, or what is causally normal in contexts where culmination is taken for granted (see also Nadathur & Filip 2021). Looking ahead, we anticipate that the causal approach can be combined with a uniform partitive theory of aspects (e.g., Altshuler 2014) to account for non-culminating uses of telic predicates in (non-)progressive contexts (Martin, 2019, a.o.).

The approach also offers an expanded view of the role of causal information in (lexical) semantic judgements. Where previous studies focused on instances of token causation (such as the use of causative verbs; Nadathur & Lauer 2020, Baglini & Bar-Asher Siegal 2021), we demonstrate that type-level information also factors into the interpretation of (non-causative) predicates, insofar as it here determines the felicity and truth conditions of telic progressives.

References

Altshuler, D. 2014. A typology of partitive aspectual operators. *Natural Language and Linguistic Theory* 32, 735–775.

Asher, N. 1992. A default, truth conditional semantics for the progresive. *Linguistics and Philosophy* 15, 463–508.

Bach, E. 1986. The algebra of events. *Linguistics and Philosophy* 9, 5–16.

Baglini, R. & Bar-Asher Siegal, E. 2021. Modelling linguistic causation, manuscript, Aarhus University and Hebrew University of Jerusalem.

Bonomi, A. 1997. The progressive and the structure of events. *Journal of Semantics* 14, 173–205.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

Landman, F. 1992. The progressive. *Journal of Semantics* 1, 120–124.

Martin, F. 2019. Non-culminating accomplishments. *Language and Linguistics Compass* 13, 1–20.

Martin, F. & Schäfer, F. 2012. The modality of *offer* and other defeasible causatives. In N. Arnett & R. Bennett (eds.), *Proceedings of the West Coast Conference on Formal Linguistics 30*, pp. 248–58, Somerville, MA: Cascadilla Press.

Nadathur, P. & Filip, H. 2021. Telicity, teleological modality, and (non-)culmination. In *Proceedings of the West Coast Conference on Formal Linguistics*, vol.̃39, p. to appear.

Nadathur, P. & Lauer, S. 2020. Causal necessity, causal sufficiency, and the implications of causative verbs. *Glossa,* 5, 49.1–37.

Parsons, T. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*, vol.̃19 of *Current Studies in Linguistics*. Cambridge, MA: MIT Press.

Pearl, J. 2000. *Causality*. Cambridge: Cambridge University Press.

Schulz, K. 2011. If you'd wiggled A, then B would've changed. *Synthese* 179, 239–251.

Szabó, Z. G. 2008. Things in progress. *Philosophical Perspectives* 22, 499–525.

Varasdi, K. 2014. Making progressives: necessary conditions are sufficient. *Journal of Semantics* 31, 179–207.

Intentional omissions are activities

Kaisa Kärki
Postdoctoral fellow
Practical Philosophy
University of Helsinki

Intentional omissions are a special group of an agent's omissions. They happen out of the agent's will, instead of being merely things that happen *to* them. I can intentionally omit to vote in an election, omit to answer a question at a dinner, or omit to mow the lawn throughout the summer. What is intentional in an intentional omission is what the agent does *not* do. Although intentional omissions are something we deemed to responsible of, they are not intentional actions because no intentional bodily movement of the agent is necessarily involved (Clarke 2010; 2014).

A good ontological account of intentional omissions needs to account for *how* they reside in space and time. In the following, it is argued that when an agent is intentionally not performing an action, there is something going on, and this something is best described as a process of a kind in which the agent is instigating, controlling, and sustaining an omission of hers. This is because intentional omissions have several features of processes; they are homogenous, continuous, unbounded, indefinite, and directly uncountable. As processes instigated and sustained by *humans*, they should be seen as activities.

Originally, a distinction between performances and activities was based on Kenny's (1963) and Vendler's (1957) analysis. They argued that differences in verb aspect between performance and activity verbs mirror differences in the way these occurrences essentially reside in time. In the following, I apply this distinction and later findings on the necessary features of processes to the metaphysics of intentional omissions. It must be noted that the data we have of the metaphysics of agency includes linguistic evidence as well intuitions, thought experiments, and phenomenological data. I assume that agents are, to some extent, experts when it comes to distinctly *human-induced* occurrences.

(1) Kenny originally distinguished static verbs such as 'know' and 'be happy' from continuous verbs such as 'learn' or 'look for' (1963: 172). He further divided continuous verbs into performance verbs such as 'kill' and 'decide whether' and activity verbs such as 'keep a secret' or 'live at Rome' (1963: 173). According to Kenny, there is an essential difference in *how* these occurrences are in time: whereas states may last for a time, performances take time and activities go on for a time (Kenny 1963: 176). Verbs that are commonly used to describe intentional omissions behave like Kenny's activity verbs. It is not plausible to say, 'I took me all summer to not mow the lawn' but we can correctly say 'I have been refraining from smoking for a decade'. One can say that answering a question took two minutes, but we cannot say that not answering a question took the whole afternoon. Not answering a question does not itself *take* time, but it can go on for a certain time-frame. Intentionally refraining from working during a strike, for instance, can go on for two weeks.

(2) Another way to distinguish performances from activities, according to Kenny, is that whereas performance verbs can happen *quickly or slowly*, activity verbs cannot (1963: 176-177). Expressions of, as well as intuitions about intentional omissions, function like activities in this way as well. One cannot refrain from smoking quickly or slowly whereas one can smoke a cigarette slowly or rapidly. Intentionally omitting from working cannot happen quickly or slowly whereas performing an action that can be completed, such as bringing a pizza home, can happen slowly or rapidly.

(3) Compared to performances, activities, and processes, are *homogenous*. This means that what is going on in a process has the same nature throughout the time-frame in which it is going on (Mourelatos 1978: 416). Any part of the process has been deemed to be of the same nature as the whole (Vendler 1957: 146). Intentional omissions are process-like in this sense as well because they have the same nature throughout the time-frame in which they are going on: there is no difference in not answering a question at the beginning or the end stages of the omission. Compared to actions, there are different

parts in answering a question, for instance, and the action is not of the same nature at every moment of its course.

(4) Another feature of processes is that they are *continuous* compared to events. Whereas an event does not exist entirely at any time during its course (Stout 1997: 25), what is going on in a process is continuously present in its entirety at different times (Stout 1997: 26). What is going on in an intentional omission, as well, is something continuous rather than a specific, concrete change or a set of changes. An agent intentionally not mowing the lawn contributes to the same continuous omission that exists entirely throughout every small decision not to mow the lawn. What is unfolding exists entirely at any specific time during the course of the omission as there is no change that requires different *stages* of an event.

(5) Intentional omissions are also *unbounded*. Whereas events are deemed to be bounded – they have a definite duration – processes endure unbounded in time (Galton and Mizoguchi 2009: 4-5). Intentionally not answering a question, for instance, is unbounded in a sense that its temporal boundaries are fuzzy. In intentionally not answering a question, there are moments from which it cannot be determined whether intentionally not answering a question has started or is still going on although there are moments from which we can definitely say that the intentional omission is unfolding.

(6) Processes also involve no culmination of an anticipated result (Mourelatos 1978: 204). Michael Bennett says that activities are represented by *open intervals* whereas performances are represented by *closed intervals* (1977: 505.) This feature has been called the *indefiniteness* of the time stretch of activities (Mourelatos 1978: 204). According to Vendler, activities such as running or pushing a cart have no terminal set point or climax (1957: 145). Activities therefore have been seen as being essentially *atelic*, that is, processes such as pushing a cart qualify as activities regardless of whether the cart is pushed to a certain end point or not, or whether the activity is goal-directed or not (Mourelatos 1993: 386). Performances, however, are *telic*, that is, in them, the end point gives closure to what was going on (Mourelatos 1993: 386). How we perceive our intentional omissions suggests that they are activities in this aspect as well. Intentional omissions are anticlimactic. One can intentionally omit to smoke, but the intentional omission never reaches an end point after which 'the deed was done.' Not taking up a topic at a meeting ceases when the meeting is over, but the end point of the activity is determined by external reasons, the activity itself does not reach a culmination point. This is also in part revealed by how we speak of our intentional omissions. The question, 'How long did you omit to pull the weeds?' is appropriate whereas there is something wrong with the question, 'How long did it take for you to not vote?' The latter kinds of descriptions are used for performances, whereas the first kinds are used to talk about activities (Vendler 1957:145). This difference is based on the different way that activities and performances endure in time – performances take a definite time because activities go on for an indefinite timespan *without* a culmination point (Vendler 1957: 145).

(7) Intentional omissions are also *directly uncountable*. Whereas it is possible to count events, processes cannot be counted — at least in the same way as events can (Galton and Mizoguchi 2009: 4). John's not smoking at a party is not directly countable, whereas George's smoking happened three times. Instead, processes are *measured* – they are individuated by extrinsic containers (Mourelatos 1978: 210). One's intentional omission to not take up a topic at a meeting can be measured extrinsically as lasting throughout the meeting. But it is difficult to perceive not taking up a topic happening three times because no specific change corresponding to this intention of the agent actually happened at the meeting.

References

Clarke R. (2010) 'Intentional omissions'. Noûs, 44 (1), 158-77.

Clarke, R. (2014) Omissions. Agency, Metaphysics, and Responsibility. New York: Oxford University Press.

Galton, A. & Mizoguchi, R. (2009) 'The water falls but the waterfall does not fall: new perspectives on objects, processes, and events'. Applied Ontology, 0, 1-33.

Kenny, A. (1963) Action, Emotion, and Will. Studies in Philosophical Psychology. New York: The Humanities Press.
Mourelatos, A. P. (1993) ‘Aristotle’s kinêsis/energeia distinction: a marginal note on Kathleen Gill’s paper’. Canadian Journal of Philosophy, 23 (3), 385-88.
Mourelatos, A. P. D. (1978) ‘Events, processes, and states’. Linguistics and Philosophy, 2, 415-34.
Vendler, Z. (1957) ‘Verbs and times’. The Philosophical Review, 66(2), 143-60.

WIELDING AND LOSING CONTROL: RUSSIAN PERFECTIVE PERFORMATIVES IN INTERACTION

Marco Biasio (Università degli Studi di Padova)

Action and control. From the 1950s onward, a growing body of sociological and psychological research has explored several different constructs related to the multifaceted notion of 'control', which has accordingly been given distinct and sometimes conflicting definitions. One fundamental theoretical distinction, as elaborated on, a.o., in SKINNER (1996), has been drawn between 'objective control' (i.e., the necessarily causal relationship between a source response or stimulus and a desired outcome) and 'subjective' or 'perceived control', which has rather a phenomenological character (i.e., fed by the subject's relative and ever-changing perception of reality). Linguistic applications of the notion(s) of 'control', which started gaining a foothold in the 1970s, have played a pivotal role both in setting up formal theories of action (i.e., BRENNENSTUHL 1982: 7–24) and in grammatical theory, where a basic opposition between the semantic categories of 'control' and 'non-control' (and their morphosyntactic reflections thereof) has been sporadically assumed as the cornerstone of some natural semantic taxonomies of lexical predicates (KLAIMAN 1991). Likewise, in Soviet (then Russian) linguistics an intuitive notion of 'control' has been a central classificatory criterion (BULYGINA 1982: 68–82) with respect to the well-formedness of several Russian morphosyntactic constructions (a.o., preventives, prescriptive infinitives, dative-infinitive constructions, and aspectual imperfective (IPF) futures in egophoric utterances). Non-binary definitions of 'control' have been often proposed against the background of the speaker's 𝕊 contextual intentions (ZALIZNJAK 1992: 63–64), concerning both 𝕊's goal-oriented mental operations and physical energies ('intensional' vs. 'denotational' control in KUSTOVA 1992: 145–146). More recently, a tighter connection between 𝕊's (contextually salient) objective or perceived control and the abstract temporal schemata underlying aspectual morphology has been put forward as well; special reference is being made to the constraints on the licenseability of perfective (PF) forms in particular syntactic environments, such as explicit performative utterances, a particular kind of self-referential declaratives which are assigned a different illocutionary force in the syntax (GROENENDIJK, STOKHOF 1976; SEARLE 1989).

The puzzle. Russian performative verbs in egophoric utterances typically come as IPF present forms (e.g., *prošu* 'I ask for'). PF present forms (e.g., *poprošu* 'I ask for') can be licensed as well insofar as a certain number of preconditions at the syntax-pragmatics interface are met, a.o., the availability and salience of 𝕊's control in the speech act event. This can be derived if we assume that Russian PF entertains two basic pragmatic components, i.e., an assertion that the corresponding action has reached its end-point and a corresponding inference (generated as a scalar implicature) that the action has started (GONCHAROV 2020: 58). Interestingly, the vast majority of Russian unpreverbed performatives enter their aspectual pair with a PF predicate obtained via morphological attachment of the aspectual operator PO-. A cluster of such performatives, conveniently enlisted on the basis of the illocutionary act they formally realize (i.e., declarations, commissives, and directives), is given in (1 a–c);

(1) a. *ka-ja-t'-sja*^{IPF}/***PO-ka-ja-t'-sja***^{PF} 'to confess', 'to repent'
b. *klj-a-s-t'-sja*^{IPF}/***PO-klj-a-s-t'-sja***^{PF} 'to swear';
ruč-a-t'-sja^{IPF}/***PO-ruč-i-t'-sja***^{PF} 'to assure';
spor-i-t'^{IPF}/***PO-spor-i-t'***^{PF} 'to bet';
c. *pros-i-t'*^{IPF}/***PO-pros-i-t'***^{PF} 'to ask (for)';
sovet-ova-t'^{IPF}/***PO-sovet-ova-t'***^{PF} 'to suggest';
xodatajstv-ova-t'^{IPF}/***PO-xodatajstv-ova-t'***^{PF} 'to solicit';
treb-ova-t'^{IPF}/***PO-treb-ova-t'***^{PF} 'to demand'

Likewise, the same pattern happens to be productive for delocutives (2 a–b) and unpreverbed biaspectual performatives which still derive preverbed PF counterparts (3 a–b);

(2) a. *blagodar-i-t'*^{IPF}/***PO-blagodar-i-t'***^{PF} 'to thank';
b. *žel-a-t'*^{IPF}/***PO-žel-a-t'***^{PF} 'to wish';
(3) a. *obeščа-a-t'*^{BA}/***PO-obeščа-a-t'***^{PF} 'to promise';
b. *vel-e-t'*^{BA}/***PO-vel-e-t'***^{PF} (*/po-vel-e-v-a-t'*^{IPF}) 'to order'

In the abovementioned examples the aspectual operator PO- is endowed with a complex semantics; in addition to introducing an abstract temporal function alongside which the given event is measured out, thus providing the event with a temporal bounding in the manner of a VP-external preverb, it may additionally signal that a certain limit has been contextually achieved, thus functioning as a proper VP-internal telicizer. The proposed semantics for PO- is tentatively spelled out in (4) below;

(4) $[\![\text{PO-}]\!]$ =
a. $[\![\text{PO-}_{\text{DEL}}]\!] = \lambda P.\lambda t.\exists e[t \supseteq \tau(e) \wedge H(P)(e)] \oplus$
b. $[\![\text{PO-}_{\text{RES}}]\!] = \lambda P.\lambda t.\exists e[P(e) \wedge \tau(e) \subseteq t]$

The proposal. In this talk I would like to propose that the selection of PO- as a perfectivizing operator for unprefixed Russian performatives is motivated by independent semantic, syntactic, and pragmatic reasons. More specifically, the main claim of the present contribution is that PO- functions as a specific piece of inflection spelling out lower in the structure a Perceived Control variable which is generated in the higher portion of the clausal spine (i.e., in the SpeechActP shell), relativized to 𝕊 (which is also the syntactic subject) and therein mediated by a λ-operator (cf. more technical details in PORTNER, PAK, ZANUTTINI 2019; ZU 2018: 101–102). Thus, the proposal aims at showing that 𝕊's intention-based perceived control is indeed grammatically relevant in Russian. This claim is substantiated by relying on the following pieces of evidence:

Syntax-semantics interface: Historical data seemingly suggest that the "double-access" PO- in (4) evolved as a further contextual variant of a new delimitative reading assigned to the preverb from the 17th century onward, i.e., during a period of significant structural changes towards the grammaticalization of the East Slavic aspectual system (DICKEY 2007) which also included the later restructuring of the allocutive forms of Middle Russian pronominal referential system and the stable introduction of a grammaticalized T-V distinction. It is thus proposed that the temporal bounding imposed on the speech act event by 𝕊 themselves can be contextually reinterpreted as a token of 𝕊's (Perceived) Control, which leads to an altering (either a strenghtening or a softening) of the intensity of the illocutionary force IF assigned to the speech act. Some examples in isolation (see 5 below) are indeed ambiguous between a more polite (↓ IF) and a more authoritarian reading (↑ IF) – both pragmatic effects being brought about by the structural relevance of 𝕊's (Perceived) Control;

(5) ***PO****-proš-u* *vaš-i* *bilet-y*.
ASK FOR.Pres.PF.1.sg. your.acc.pl. ticket.acc.pl.
'Tickets(!)' (lit. 'I ask for your tickets')

Pragmatics: Interface approaches to the structural mapping of discourse roles such as ZU (2018), however, have been criticized for being too rigid and substantially misrepresenting the dynamic and ever-changing nature of the speech act event, including the mutual accomodation of new propositions into the common ground (STALNAKER 2014) and the active role of the hearer ℍ in (re)shaping context (WILTSCHKO 2021). Following BENZ's (2021) assumption that conversations are defined by sequences of joint coordinated actions between 𝕊 and ℍ (the so-called 'joint projects') constrained by epistemic maxims of licensing and uniqueness, it is therefore proposed that the contextual felicity and appropriateness of PF PO-performatives is being constantly evaluated and renegotiated against a set of pragmatic and conversational variables, a.o., the role and position of 𝕊 towards ℍ (viz. their ascendency over ℍ), 𝕊's perceived control over the eventuality, and the successful update of both 𝕊 and ℍ's information states. Consider the following complex communicative situation (6).

[*Context*: You (i.e., '𝕊') have been employed as an office worker for the same company for the last twenty years. For some unclear reasons a new young colleague of yours, let him be called Vladimir Vladimirovič (i.e., 'ℍ'), has been scoffing you at work for some time now. Tired of cracking light-hearted smiles in response to his constant banter, one day you eventually snapped and confronted him directly:]

(6) **𝕊**: Vladimir Vladimirovič, ja ser'ezno govorju.
Vladimir.nom. Vladimirovič.nom. I.nom. seriously SPEAK.Pres.IPF.1.sg.
Ja ot Vas ***PO****-treb-u-ju*, čtoby Vy ko
I.nom. from you.acc.pl. DEMAND.Pres.PF.1.sg. COMP you.nom.pl. towards
mne projavljali uvaženie!
I.dat. SHOW.Past.IPF.m.pl. respect.n.acc.sg.
'Vladimir Vladimirovič, I am dead serious now. I demand that you show respect for me!'
ℍ: A kto Vy takoj, čtoby
and who.nom. you.nom.pl. such.m.nom.sg. COMP
otdavat' takie prikazy?
ISSUE.Inf.IPF such.m.acc.pl. order.m.acc.pl.
'And who are you to be issuing such orders?'
𝕊: Ja tol'ko *treb-u-ju*, čtoby Vy prekratili
I.nom. only DEMAND.Pres.IPF.1.sg. COMP you.nom.pl. STOP.Past.PF.m.pl.
izdevatel'stvo nado mnoj, vot i vse.
mockery.n.acc.sing. over I.inst. PART and all.n.nom.sg.
'I only demand / I am only demanding that you stop laughing at me, that's it'

The aspectual switch PF (***PO****-treb-u-ju*) → IPF (*treb-u-ju*) in the same joint project can be parsed in two possible ways; either 𝕊 wrongly believes they can bring about a desired outcome by exerting control over the speech act event (PF is infelicitous), or ℍ unexpectedly refuses to accomodate 𝕊's legit demands into the common ground (control encoding crashes and PO- is not spelled out). Following the definition of a context *c* <$\mathcal{P}$, cs> given in PORTNER, PAK, ZANUTTINI (2019: 16), the relation can be modelled as follows (7 a–c);

(7) 𝕊: A: c ⟨P, cs⟩ ^ P = h : h(P_1) = {N_1} and h(P_2) = {N_1} (𝕊's perceived control – demand towards ℍ)
ℍ: c ⟨P, cs⟩ ^ P = h : h(P_1) = {N_1} and h(P_2) = {N_1, N_2} (ℍ challenges 𝕊's control)
𝕊: c ⟨P, cs⟩ ^ h : h(P_1) = {N_1, N_2} and h(P_2) = {N_2} (𝕊 retreats – no spell-out of [control])

Selected references Benz 2021. Benz A., *Epistemic Perspectives and Communicative Acts*, in Frontiers in Communication 6(612733), pp. 1–20. DOI: 10.3389/fcomm.2021.612733. • Brennenstuhl 1982. Brennenstuhl W., *Control and Ability: Towards a Biocybernetics of Language*. Amsterdam-Philadelphia: John Benjamins. • Bulygina 1982. Bulygina T.V., *K postroeniju tipologii predikata v russkom jazyke*, in Seliverstova O.N. (ed.), Semantičeskie tipy predikatov. Moskva: Nauka, pp. 7–85. • Dickey 2007. Dickey S., *A Prototype Account of the Development of Delimitative* po- *in Russian*, in Divjak D., Kochańska A. (eds.), Cognitive Paths into the Slavic Domain. Berlin-New York: Mouton de Gruyter, pp. 329–374. DOI: 10.1515/9783110198799.4.329. • Goncharov 2020. Goncharov J., *Whom to oblige?*, in Marušič L., Mišmaš P., Žaucer R. (eds.), Advances in Formal Slavic Linguistics 2017. Berlin: Language Science Press, pp. 51–73. DOI: 10.5281/zenodo.3764847. • Groenendijk, Stokhof 1976. Groenendijk J., Stokhof M., *Some Aspects of the Semantics and Pragmatics of Performative Sentences*, in Groenendijk J., Stokhof M. (eds.), Amsterdam Papers in Formal Grammar, Vol. I. Proceedings of the Amsterdam Colloquium on Montague Grammar and Related Topics. Amsterdam: Centrale Interfaculteit Universiteit van Amsterdam, pp. 61–94. • Klaiman 1991. Klaiman M. H., *Control and Grammar*, in Linguistics 29(4), pp. 623–651. DOI: 10.1515/ling.1991.29.4.623. • Kustova 1992. Kustova G.I., *Nekotorye problemy analiza dejstvij v terminax kontrolja*, in Arutjunova N.D., Rjabceva N.K. (eds.), Logičeskij analiz jazyka. Modeli dejstvija. Moskva: Nauka, pp. 145–150. • Portner, Pak, Zanuttini 2019. Portner P., Pak M., Zanuttini R., *The Speaker-Addressee Relation at the Syntax-Semantics Interface*, in Language 95(1), pp. 1–36. DOI: 10.1353/lan.2019.0008. • Searle 1989. Searle J. R., *How Performatives Work*, in Linguistics and Philosophy 12(5), pp. 535–558. DOI: 10.1007/BF00627773. • Skinner 1996. Skinner E. A., *A Guide to Constructs of Control*, in Journal of Personality and Social Psychology 71(3), pp. 549–570. DOI: 10.1037//0022-3514.71.3.549. • Stalnaker 2014. Stalnaker R., *Context*. Oxford: Oxford University Press. • Wiltschko 2021. Wiltschko M., *The Grammar of Interactional Language*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108693707. • Zaliznjak 1992. Zaliznjak A.A., *Issledovanija po semantike predikatov vnutrennego sostojanija*. München: Verlag Otto Sagner. • Zu 2018. Zu V., *Discourse Participants and the Structural Representation of the Context*, PhD dissertation, New York University.

Agentive Modals and Agentive Modality: A Cautionary Tale

Tim Kearl, University of Arizona

Robert H. Wallace, California Polytechnic State University

Abstract: In this paper, we consider recent attempts to account for the metaphysics of agentive modality in terms of conditional statements. We argue that there are reasons to be pessimistic about gleaning a metaphysics of agency from a conditional semantics for agentive modals like "S can ϕ" or "S cannot but ψ".

Views of the sort under discussion have an important pedigree in philosophy starting with the classical conditional analysis of free will, according to which an agent is able to do otherwise than they in fact do if, and only if, had they wanted (or chosen) to do otherwise, then they would have (Moore 1912, Ayer 1956). This classical analysis faces well-known problems; most pressingly, it fails to capture the true extension of agentive modal claims (Lehrer 1968). For instance, an agent who has a severe phobia of the color red is not able to eat a red piece of candy even though it is true that, had they wanted or chosen to do so, they would have. It's just that, if they had wanted or chosen to, then they would not have had the phobia.

We consider two recent versions of the strategy that appeals to a conditional semantics of agentive modals to get some traction on the metaphysics of agentive modality. First, we consider Mandelkern, Schulteis, and Boylan's (2017) semantics, an updated version of the classical view that they call the Act-Conditional Analysis ("ACA"). We also consider David Lewis's (2020) posthumous compatibilist analysis of abilities in terms of the absence of obstacles to action, which we term the Obstacle-Free Analysis ("OFA"). The former is specifically billed as a semantics for agentive modal claims in natural language, whereas the latter is explicitly a metaphysician's analysis of abilities. It is important to note that we are not pessimistic about any semantics specifically, but rather the methodology. In particular, one shouldn't try to settle the metaphysics of agentive modality by way of a conditional semantics of agentive modals.

We suggest that although the ACA and OFA are significant improvements on the classical conditional analysis, they face counterexamples of two sorts, depending on whether the conditionals are construed subjectively—their truth depending on mental states of the agent, or perhaps what is within the agent's deliberative reach—or objectively—their truth depending on all the relevant facts, whether or not those facts are beyond the agent's ken. Our diagnosis of the counterexamples is that both the ACA and the OFA sin in taking some *basic* agentive modality for granted. In particular, the abilities to perform basic actions, or "basic abilities", feature as a primitive in these theories. Here is where the semantic and metaphysical projects most clearly are at odds; while it is perfectly acceptable for a semantics of agentive modal claims to take some modality for granted in getting the extension of action claims correct, a metaphysical explanation of agentive modality cannot, at least not in the way that these conditional approaches to agentive modality do.

Because the classical conditional analyses of abilities had this problem just as much as do their contemporary metaphysical cousins, dispositional analyses of abilities (Fara 2008, Vihvelin 2013), we conclude by a pessimistic induction that (probably) no conditional approach to agentive modality will succeed.

This leaves open a strange and potentially disturbing consequence of our argument. Namely, the semantics and metaphysics of agentive modality may come apart at the seams. We briefly consider the merits of Vetter's (2013) view, one which leans into that consequence at the cost of robust anti-reductionism about agentive modality. Her semantics starts from the idea that agentive abilities–basic abilities among them–explain the truth of agentive modal claims. Unfortunately, this way of connecting the semantics and the metaphysics of agentive modality is antithetical to the longstanding enterprise of gleaning, more or less reductively, a metaphysics of agentive modals from its best semantics.

References

Ayer, A. J. 1954. "Freedom and Necessity." *Philosophical Essays*. New York: St. Martin's Press: 3-20.

Fara, M. 2008. "Masked Abilities and Compatibilism." *Mind* 117 (468): 843-65.

Lehrer, K. 1968. "Can's Without 'If's." *Analysis* 24: 159-60.

Lewis, D.. 2020. "Outline of Nihil Obstat: An Analysis of Ability". *The Monist* 103: 241-244

Mandelkern, M., G. Schultheis, and D. Boylan. 2017. "Agentive Modals." *Philosophical Review* 126 (3): 301-343.

Moore, G. E. 1912. *Ethics*. Oxford: Oxford University Press.

Vihvelin, K.. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford University Press.

Vetter, B. 2013. "Semantics for Anti-Humeans" *Philosopher's Imprint* 13 (16): 1-27

Analyzing *intentionally* with local listening and would-be preventers

Bridget Copley and Clémentine Raffy, SFL (CNRS/Paris 8)

The Knobe effect: What does it mean to do something intentionally? (1a) is often reported false but (1b) is often reported true, in contexts where where helping/harming the environment is a side-effect of the chairman's action, even though the chairman says in both cases she doesn't care about helping/harming the environment (Knobe 2003; the "Knobe effect").

(1) a. Help condition: "The chairman intentionally helped the environment."
b. Harm condition: "The chairman intentionally harmed the environment."

Our contribution here is twofold. We propose that causal models with ***local listening***, where each edge (arrow) is associated with its own dependency function on truth values (Copley, to appear), can shed light on the Knobe effect. The idea is that the only functions that can be associated with arrows are those where the influenced truth value actually depends on the influencing truth value; otherwise there would be no influence (this is "listening" as in Pearl 2000; here it is "local" to each arrow). Another contribution is the proposal that the meaning of *intentionally* relies on the agent being a ***would-be preventer*** (McGrath 2005).

New data: In support of this idea, note that French *laisser* 'let' requires would-be preventer subjects (Raffy 2021). In this it contrasts with English *let*, which does not have this requirement. Given the chairman scenario, (2a), corresponding to the Help condition, is odd, while (2b), corresponding to the Harm condition, is felicitous.

(2) a. ??Le PDG a laissé les employés améliorer l'environnement.
the chairman AUX let the employees better the-environment
'The chairman let the employees help the environment.'
b. Le PDG a laissé les employés nuire à l'environnement.
the chairman AUX let the employees harm to the-environment
'The chairman let the employees harm the environment.'

Local listening: Classically in causal models (e.g. (3a)) the value of an endogenous variable Y is given by a function on all the variables that Y depends on, as shown in (3b). Sloman et al. (2012) use causal models fruitfully in an analysis of the Knobe effect. However, they use probabilities as the values of the variables, which is not useful for (most) formal semantic approaches. An approach using truth values can, however, have similar flexibility: Following Copley 2021, we alter the framework such that each arrow corresponds to its own function, representing the dependency that occurs if all other nodes are erased ("if all else is equal"). Where conflicts arise, an otherwise expected influence can be blocked from determining the value of the endogenous variable, which we notate using a double-barred arrow: $X \nrightarrow Y$.

(3) a. $X \rightarrow Y \leftarrow Z$ b. $F(X, Z) = Y$ c. $F(X) = Y$ and/or $F'(Z) = Y$

Indifference and disjunctive values We assume a third truth value "indiff" representing indifference, for nodes representing desires such as those of the CEO. We also allow for the returned value of an arrow function to be a disjunction between two truth values; such a disjunction licenses either of its values for the node in question.

(4) a. **Meaning of *intentionally*:** Let $D_{@p}$ represent an desire toward either p or $\neg p$. *x intentionally p* presupposes the model in (5), and is true iff $D_{@p}$ is a would-be-preventer for p.
b. $D_{@p}$ is a would-be-preventer for p iff there is a path via arrows from $D_{@p}$ to R_p with all values licensed, and there is a possible value of $D_{@p}$ that licenses $R_p = 0$.

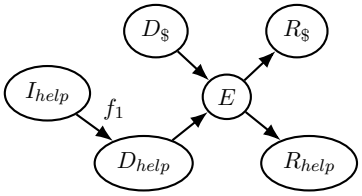
(5)
$$\underbrace{\underset{\text{whether desire-@p}(x)}{D_{@p}} \xrightarrow[influences]{}}_{\text{contributed by } \textit{intentionally}} \quad \underbrace{\underset{\text{whether } \exists e : \text{agent}(x, e)}{E} \xrightarrow[influences]{} \underset{\text{whether } \exists e' : p(e)}{R_p}}_{\text{contributed by rest of sentence}}$$

In words: whether the result occurs depends causally on whether the agent's action occurs (this is the not-at-issue meaning), and whether the agent's action occurs depends causally on whether the agent is a has an intention about the result and could have an intention to prevent the result (this is the at-issue meaning).

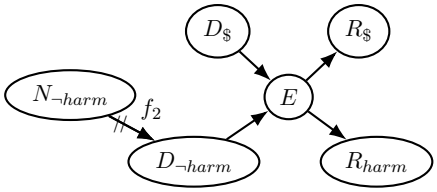
Models for the Help and Harm conditions: Along the lines of the above discussion, we argue for the functions below for ideals (e.g. I_{help}) and for norms (e.g. $N_{\neg harm}$). Having an ideal ($I_{help} = 1$) doesn't require you to want to realize it, while holding to a norm ($D_{\neg harm} = 1$) does.

(6)

a. Help condition



b. Harm condition



Function f_1 associated with $I_{help} \rightarrow D_{help}$ in (6a)

I_{help}	D_{help}
1	1∨indiff
0	0

Function f_2 associated with $N_{\neg harm} \rightarrow D_{\neg harm}$ in (6b):

$N_{\neg harm}$	$D_{\neg harm}$
1	1
0	$1 \vee 0$

What decides the judgments, according to (5), is whether there is a licensed line of the truth table from the desire $D_{help/\neg harm}$ to E such that $E = 0$. If D is not influenced by anything and can freely choose, then the CEO is a would-be preventer. However, if D is influenced by another node, it may not allow for such a line in the table. We assume that if the context doesn't make us block the D to E influence, it remains in the model. Because the actual value of I_{help} licenses the actual value of D_{help} (namely, indiff; see f_1), we don't block the arrow between those nodes, and the actual value of I_{help} does not permit $D_{help} = 0$, so the CEO cannot be a would-be preventer, and (1a) is false. But because $N_{\neg harm}$ does not allow $D_{\neg harm} =$ indiff (see f_2), we have to block the $D_{\neg harm}$ to E influence. This blocking allows $D_{\neg harm}$ to counterfactually have the value 0 and thereby make $E = 0$, making the CEO a would-be preventer and (1b) true.

We will further show how this analysis works for Machery's (2008) "Smoothie" scenario. The agent is not a would-be preventer in (7a), in the version where the action is less typically judged intentional, but is one in (7b), in the version where the action is more typically judged intentional.

(7) a. ??Le client a laissé l'employé lui donner une tasse commémorative.
the customer AUX let the-employee him give a cup commemorative
'The customer let the employee give him a commemorative cup.'

b. Le client a laissé l'employé lui faire payer 1 dollar de plus.
The customer AUX let the-employee him make pay 1 dollar of more
'The customer let the employee charge him a dollar extra.'

References: Copley, B., to appear. Reconciling causal and modal representations for two Salish out of control forms. WCCFL 39. **Knobe, J. 2003.** Intentional action and side effects in ordinary language. *Analysis*, 63, 190–93. **Machery, E., 2008.** The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language,* 23(2), 165-189. **McGrath, S., 2005.** Causation by omission: A dilemma. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* 123(1/2), 125-148. **Raffy, C. 2021.** Letting *in Romance.* PhD diss., Univ. Paris 8 & Univ. zu Köln. **Sloman, S.A., Fernbach, P.M. and Ewing, S., 2012.** A causal model of intentionality judgment. *Mind & Language,* 27(2), 154-180.

The Group Knobe Effect Revisited. Epistemic and Doxastic Side-Effect Effects in Intuitive Judgments Concerning Group Agents.

Maciej Tarnowski, Adrian Ziółkowski, Mieszko Tałasiewicz
m.tarnowski3@student.uw.edu.pl
adrian.a.ziolkowski@uw.edu.pl
m.talasiewicz@uw.edu.pl
Faculty of Philosophy, University of Warsaw

Can groups perform actions and take responsibility for their consequences? And if so, in what sense? Apart from being described as acting and responsible, can a group be also described as possessing beliefs, intentions and desires? These questions are widely debated in social philosophy and metaphysics, and one's answer to them determines their philosophical position in the debate on collective agency and intentionality. Many arguments made for and against the view which ascribes groups the potential of being intentional agents and holders of intentional states and, thus, treats them similarly to individual agents (which we label "realism" about collective agency) relied on the perceived intuitiveness of such a view among the laypeople. While realists argue that the widespread use of statements like 'The court *finds* the accused guilty…' or 'Amazon *plans* to cut its employment…' presupposes their intuitive truth (e.g. Tollefsen 2002), irrealists argue that collective propositional attitudes and agents are "spooky entities" created by "magic", and that such statements are, at best, metaphorical (e.g. Thomasson 2019). Another important controversy is, if realism is correct, does it stem from collectivist or distributivist intuitions regarding group agents? According to distributivism, possession of intentional states by a group agent A is reducible to the possession of relevant states by the members of A (group-*qua*-its-members), while according to collectivism, group intentional states are irreducible (group-*qua*-group). The question whether realism or irrealism is intuitive and presupposed by folk psychology – and which is revisionary – remains not settled and taking into account the role of perceived intuitiveness in philosophical argumentation, there is a need for empirical investigation into this problem.

In our talk, we will take a closer look at the effect described in the literature as the Group Knobe Effect (GKE), which, to put it shortly, is an asymmetry in ascription of intentionality of an action performed by a group agent depending on its negative or positive side-effects. One could say it is an extension of the "regular" Knobe Effect (or the side-effect effect), which is a well-documented phenomenon noticeable in folk judgments regarding agency of individual agents. The expected asymmetry (GKE) is supposed to reveal realist intuitions. If laypersons perceive group agents similarly to individual agents when it comes to the ascription of intentionality or responsibility for side-effects of their actions, a strong abductive argument for the claim that the folk tend to hold realist intuitions about group intentionality and responsibility is available. The hypothesis that the folk are realists about group intentional action is simply "the best explanation" of the existence of GKE (at least until the philosophers who claim that realism is counterintuitive come up with an

alternative explanation of the existence of GKE). The claim that the individual Knobe Effect obtains only with respect to intentional agents is assumed by various analyses of KE, which explain this effect by the folk-psychological mechanisms of belief attribution or characteristics of the folk concept of intentional action (see e.g.: Knobe 2006, Feltz 2007, Alfano, Beebe, Robinson 2012, Paprzycka-Hausman 2020). Some studies have shown that the asymmetry may be also observed in attributions of knowledge (Beebe, Buckwalter 2010, Beebe, Jensen 2012) and belief (Beebe 2013), which also supports this thesis. Therefore, if a similar asymmetry is observed with respect to group action, it would provide evidence that groups are intuitively taken by the folk to be intentional agents.

We will present the results of two experiments regarding GKE we conducted. Our experiments successfully replicated the findings reported by Michael and Szigeti (2019) who first observed the Group Knobe Effect in folk judgments concerning intentionality of action and moral responsibility. We also found empirical evidence of the existence of two related effects: the Group Epistemic and Doxastic Knobe Effects (GEKE and GDKE), which show analogous asymmetry in folk judgments with respect to knowledge and belief ascriptions to groups. Observing these two effects further strengthens the claim that laypeople perceive groups as intentional agents as well as provides evidence for the claim that groups are perceived as knowers and believers in a way analogous to individuals. In our detailed analysis of the data, we will also address the issue whether the apparent realist intuitions we observed stem from a collective or distributive perception of group agents. We will argue that the empirical material available thus far does not allow to say that either of these views is common among laypersons: it rather seems that there individual differences in exhibiting collectivist and distributivist intuitions concerning group agency. In our talk, we hope to explain how the existence of the Group Knobe Effect and its epistemic and doxastic counterparts impacts the philosophical debate on collective agency and intentionality and supports the claim about the intuitiveness of realism regarding collective agency among the laypeople.

References:

1 Alfano, M., Beebe, J. R., Robinson, B. (2012). The centrality of belief and reflection in Knobe-effect cases: A unified account of the data. *The Monist*, *95*(2), 264-289.

2 Beebe, J. R. (2013). A Knobe effect for belief ascriptions. *Review of Philosophy and Psychology*, *4*(2), 235-258.

3 Beebe, J. R., Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, 25(4): 474-498.

4 Feltz, A. (2007). The Knobe effect: A brief overview. *The Journal of Mind and Behavior*, 265-277.

5 Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3): 190-194.

6 Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical studies*, *130*(2), 203-231.

7 Michael, J. A., Szigeti, A. (2019). "The Group Knobe Effect": evidence that people intuitively attribute agency and responsibility to groups. *Philosophical Explorations*, *22*(1), 44-61.

8 Paprzycka-Hausman, K. (2020). Knowledge of consequences: an explanation of the epistemic side-effect effect. *Synthese*, *197*(12), 5457-5490.

9 Thomasson, A. L. (2019). The ontology of social groups. *Synthese*, *196*(12), 4829-4845.

10 Tollefsen, D. (2002). Organizations as True Believers, *Journal of Social Philosophy*, 33(3): 395–401.

Analytical lexical strategies denoting agency and causativity.
A synchronic and diachronic investigation in Italian and other Romance languages

Valentina Piunno
(Roma Tre University)

Vittorio Ganfi
(University of Modena and Reggio Emilia)

Despite the growing interest in the notions of agency and causativity (Cruse 1973; Shibatani 1976; Comrie 1976; DeLancey 1984; Song 1996; Dixon 2000; Talmy 2000; Haspelmath & Müller-Bardey 2001; Wolff & Song 2003; Lehmann 2017), the contribution of lexical (analytical) strategies concerning the use of complex prepositions to the designation of agentive and causative relations is still neglected. In particular, this paper aims at analyzing the set of Agentive Complex Prepositions, i.e. fixed word combinations having a prepositional function and signaling the Agent of a predicate (e.g. *by the hands of*), as well as the causer of a causative construction (e.g. *at the request of*), cf. the following example from Italian:

(1) IT L'uomo fu ucciso **su mandato del** cugino **per mano di** un killer
the man was murdered on order of.the cousin for hand of a killer
'The man was murdered by a killer, as his cousin ordered'

While in the example (1) the Complex Preposition *su mandato di* 'on mandate of' introduces the participant who conceives and plans an action, and has a coercive power on the actual performer, the Complex Preposition *per mano di* 'by the hands of' introduces the participant who intentionally and directly performs the action, being its originator.

This investigation experiments a synchronic and diachronic cross-linguistic approach. On the one hand, the synchronic perspective aims at i) identifying different types of Complex Prepositions denoting the Agent in Italian and at comparing them to two Romance languages (i.e. Spanish and French), ii) at distinguishing them according to the degree of agency, and at iii) correlating the various Complex Prepositions to the different types of causativity that can be represented in the causativity scale. On the other hand, the diachronic investigation takes into account the evolution of Romance Complex Prepositions across different historical phases, and puts into relation synchronic restrictions with the original syntactic environment within the grammaticalization contexts.

A set of Agentive Complex Prepositions of Romance languages will be collected by means of a corpus-based investigation[1]. Then, Complex Prepositions will be classified on the basis of specific agentive semantic traits, as well as according to their meaning/function. Thus, they will be assigned a specific agentive class on the basis of the degree of control they express, and, in particular, according to the following agentive traits:

i. intentionality: the volitional involvement of the participant (Dowty 1991), who "intends to let the situation happen" (Lehmann 2017, p.36),
ii. coercive power: the authority of coercing someone to do something,
iii. monitoring: the authority of control over the performing of the action (Lehmann 2017).

At least the following classes of Agentive Complex Prepositions will be considered:

(2) IT PERFORMER
sia stato ucciso **per mano di** qualche boss mafioso
(he) be killed for hand of some boss criminal
'he was killed at the hands of some mafia boss'

(3) IT CONCEIVER
Il killer ha ucciso l'uomo **su suggerimento di** John
the killer has murdered the man on suggestion of John
'Upon suggestion of John, the killer murdered the man'

[1] The synchronic analysis will be based on the *TenTen* corpora *Italian Web 2016*, the *Spanish Web 2018*, and the *French Web 2017* (Jakubíček *et al.* 2013), which will be used to extract the examples and their contexts of occurrence. For the diachronic investigation, the following corpora and dictionaries will be used: i) for Italian, *OVI Corpus* and *MIDIA corpus*, as well as the *TLIO dictionary*; ii) for French, the *Dictionnaire du Moyen Francais* and the *Base textuelle Old FRANTEXT*; iii) for Spanish, the *CORDE corpus*.

(4) IT INSTIGATOR

i carabinieri,	**su**	**mandato di**	un magistrato	hanno sequestrato	la casa
the policemen	upon	mandate of	a magistrate	have seized	the house

'Upon the mandate of a magistrate, policemen seized the house'

While *per mano di* in (2) introduces the PERFORMER of the action (i.e. the participant who intentionally performs the action, being at the same time its originator and instigator), the CP *su suggerimento di* (3) introduces a CONCEIVER (i.e. the participant who conceives the action, but does not perform it and has no control over its development nor over someone else performing it), while the CP *su mandato di* (4) introduces the INSTIGATOR (i.e. the participant who conceives and plans an action, and has a coercive power on the actual performer).

The distinction between INSTIGATOR and CONCEIVER in terms of the presence of the trait <± coercive power> recalls the dichotomy *between direct* and *indirect causation* (Comrie, 1981). As causative constructions, Romance Agentive Complex Prepositions are able to:

(a) assign a new agentive slot to the verbal unit (Lehmann 1996),
(b) give rise to the distinction between two agentive participants in a sentence, sharing the involvement in the action, but being characterized by different agentive features.

Furthermore, there is a peculiar hierarchical relation between the different Agents (*scale of causative force*, cf. Comrie 1981). This analysis also aims at correlating the various Agentive Complex Prepositions to the different types of causativity that can be represented in the causativity scale.

References

Comrie, B. (1976). The syntax of causative constructions: cross-language similarities and divergences. In Shibatani, M. (ed.) *Syntax and semantics* 6: *the grammar of causative constructions.* Academic Press: New York, 261-312.

Comrie, B. (1981). *Language universals and linguistic typology. Syntax and morphology,* Chicago: University of Chicago Press.

Cruse, A.D. (1973). Some thoughts on agentivity. *Journal of Linguistics*, 9, 11-23.

Dixon, R.M.W. (2000), A typology of causatives: form, syntax and meaning. In Dixon, R.M.W. & Aikhenvald, A.Y. (eds.), *Changing valency. Case studies in transitivity*. Cambridge: CUP, 30-83.

DeLancey, S. (1984). Notes on agentivity and causation. *SiL*, 8, 181-213.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547-619.

Fagard, B. & De Mulder, W. (2007). La formation des prépositions complexes: grammaticalisation ou lexicalisation?, *Langue française,* 4 (156), 9-29.

Haspelmath, M., Müller-Bardey, T. (2004). Valency change. In Booij, G., Lehmann, C., Mugdan, J. (eds), *Morphology: A Handbook on Inflection and Word Formation*, Berlin: de Gruyter, vol. 2, 1130-1145.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P. & Suchomel, V. (2013). The TenTen Corpus Family. In *Proceedings of the 7th International Corpus Linguistics Conference CL 2013*, UK, July 2013, 125-127.

Lehmann, C. (1996). Control in Yucatec Maya. In François, J. (ed.), *La sémantique des relations actancielles à travers les langues*. Strasbourg: UMR du CNRS LanDisCo & Université des Sciences Humaines, 105-127.

Lehmann, C. (2017). Grammaticalization and automation. *JournaLIPP*, 5, 33-48.

Shibatani, M. (1976). The grammar of causative constructions: a conspectus. In Shibatani, M. (ed.), *Syntax and semantics* 6: *the grammar of causative constructions.* Academic Press: New York.

Song, Jae J. (1996). *Causatives and causation. A universal-typological perspective*. Harlow, Longman.

Talmy, L. (2000). *Toward a cognitive semantics*, Cambridge: MIT press.

Wolff, P., Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology*, 47(3), 276-332.

Corpora and dictionaries

Base textuelle *Old FRANTEXT*:	https://www.frantext.fr
Corpus diacrónico del español (CORDE):	https://www.rae.es/banco-de-datos/corde
Dictionnaire du Moyen Français (DMF):	http://www.atilf.fr/dmf
Opera del Vocabolario Italiano (OVI) corpus:	http://www.ovi.cnr.it/index.php/it/
TenTen corpora:	https://auth.sketchengine.eu
TLIO (Tesoro della Lingua Italiana delle Origini):	http://tlio.ovi.cnr.it/TLIO/

Learning Propositional Attitudes- Becoming an Agent by Developing 'conceptual metaphors'

Shruti Sharma
Doctoral candidate, Department of Humanities & Social Sciences
Indian Institute of Technology, Jodhpur
sshruti398@gmail.com, sharma.27@iitj.ac.in

Abstract for Agency and Intentions in Language, Second Annual Interdisciplinary Workshop, Harvey Mudd College (virtual), January 12 – 14, 2022

First-person concepts (FPC) like I, me, my, myself etc are not used as soon as a child starts learning a language; FPC comes in the language of a child after a period of development. Povinelli and colleagues (Povinelli, 2001) show that it is not prior to approximately three years that children start understanding the temporal continuation[1] of their 'self'. By four years, most children start using the word 'me' instead of their proper name. And it is also by the age of four-five that children start understanding and holding various perspectives on objects and people. (Rochat 2003) Genuine usage of propositional attitude verbs (PAV), for example: believe, hope, know, etc is good evidence to say that a child has an inner self who conceives itself as an agent. Since the genuine understanding of propositional attitudes verbs (PAV) is not seen before a certain age, it creates an impression that a child's inner-self conceived as an agent also remains absent. Genuine usage of propositional attitude (PA) implies intentional behavior. Intentional behavior requires the actions to be self-generated; it should have a causal acknowledgment that the action took place because of the propositional attitude. As Olson (2007) rightly points out that the judgments about 'intentional actions' can only be ascribed to others and made by the 'self' if there exists an awareness about them, separating them as actions done 'on purpose' as opposed to actions which are mere 'accidental'.

Given that the emergence of first-person concepts (FPC) happens late and proper understanding of propositional attitude verbs (PAV) require some time, some philosophers like Quine (1960), Davidson (1999), Dennett (1978), Olson (2007) etc, argue intentional states to be a linguistically learned process. However, others like Fodor(1975), Searle (1983) believe that these are the very structural basis of the human mind. Propositional attitude (PA) For example, 'I believe It is raining', etc (which involve PAV and FPC), are the paradigmatic case of intentional states. Those (like Searle and Fodor) who understand intentionality as a given mechanism/property of the mind explain social behavior and language learning in terms of intentionality. Therefore, it appears that they need not explain the development of language and its radical effect on one's mentality. But those who understand intentionality in degrees or at least as a complex linguistic mechanism or language-dependent model have to engage in its development and admit that it is a part of linguistic development. Even if one grants the possibility that it is a given mechanism/property of the human mind as opposed to a culturally learned mechanism/property through social training, one is compelled to give the reason as

[1] Temporal continuation of the self refers to an understanding of one's own inner self as existing in past, present, and future as a continuous being.

to why language is the only medium or how it is only in and through natural language one's inner self appears.

Whether one buys the idea that it is an already given mechanism/property of the human mind (like Searle and Fodor) or it develops with language (like Davidson, Dennett, Olson), the necessity to study the development of intentional states with the development of propositional attitude (PA) cannot be eliminated. It is so because PA (which involves genuine usage of FPC and PAV) brings out the inner self as an agent that was initially lying passive. Therefore, natural language plays an indispensable role, whether as a mere *medium* to express or as a *cause* of first-person perspective. The first section of the paper/presentation will elaborate more on the first-person perspective and its connection with the inner self' as an agent. The second section will analyze Olson's (2007) 'quotation theory' about the development of intentional states. This section also emphasizes on the necessity to entertain simple, intentional states prior to the emergence of propositional attitudes (PA). The last section explores the possibility of a correlation between PA or multiple perspectives/modes of the inner self and the development of 'conceptual metaphors' (Lakoff and Johnson 1980). It elaborates upon the possible role of learning metaphorical language and developing an alternative perspective of the 'self'.

Keywords: Propositional attitudes, Simple intentional states, Metaphorical language.

References

Davidson, Donald. 1999. "The Emergence of Thought." *Erkenntnis (1975-)* 51 (1): 7–17.

Dennett, Daniel C. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.

Fodor, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Olson, David R. 2007. "Self-Ascription of Intention: Responsibility, Obligation and Self-Control." *Synthese* 159 (2): 297–314. https://doi.org/10.1007/s11229-007-9209-2.

Povinelli, Daniel J. 2001. "The Self: Elevated in Consciousness and Extended in Time." In *The Self in Time: Developmental Perspectives*, 75–95. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Quine, W., and O. Van. 1960. *Word and Object: An Inquiry into the Linguistic Mechanisms of Objective Reference*. Word and Object: An Inquiry into the Linguistic Mechanisms of Objective Reference. Oxford, England: John Wiley.

Rochat, Philippe. 2003. "Five Levels of Self-Awareness as They Unfold Early in Life." *Consciousness and Cognition* 12 (4): 717–31. https://doi.org/10.1016/S1053-8100(03)00081-3.

Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173452.

The linguistic and the Psychological Contributions to the Knobe Effect and the Limit of the Linguistic Effect.

Masaharu Mizumoto* and Zhong Yuanyan†

Mizumoto (2018) showed that the Knobe effect, or the moral asymmetry of intentionality attribution can be observed without any vignette. He used the felicity judgments about the sentences containing "intentionally", each of which expresses either a morally good, bad, and neutral action. Participants judged the sentences expressing an agent intentionally doing a morally bad thing significantly more acceptable (correct and natural) than those expressing an agent intentionally doing a morally good thing (for example, in the case of harming/improving the environment, more than 80% as opposed to 40%). However, there he also used two Japanese counterparts of "intentionally", which showed the same moral asymmetry, with significantly sharper asymmetries than the one found for English "intentionally".

Given this linguistic diversity, one can expect that there are counterparts of "intentionally" in other languages which show patterns very different from what we found in English and Japanese. In particular, the most interesting case would be an adverb which showed the opposite pattern of moral asymmetry, such that sentences with the adverb expressing a morally good (intentional) action would be judged natural, while those expressing the a morally bad (intentional) action would be judged unnatural or ungrammatical, by the native speakers. Indeed, we found such an adverb, in Chinese.

In this paper, we will report the results of surveys with three Chinese counterparts for English "intentionally": 1. Gu yi de;故意地、2. You yi de;有意地、3. Te yi de 特意地. 1 can be understood as a Chinese counterpart of Japanese "wazato", which showed the sharpest moral asymmetry in Mizumoto (2018). 2 is the standard translation of English "intentionally". But 3 is also a translation of "intentionally", "deliberately", or "specifically" into Chinese. We conducted the analogous surveys in Chinese following Mizumoto (2018)'s method, with these three adverbs.

The result of a survey with 1 showed the pattern similar to Japanese "wazato", and the result of a survey with 2 showed the pattern similar to English "intentionally", but the result of a survey with 3 showed the opposite moral asymmetry pattern, where sentences about a morally bad action were judged "unnatural" or "wrong", while sentences about a morally good action judged mostly "natural".

The question then is how ordinary people would respond to the standard Chairman case asked using these adverbs, which we also conducted with Chinese participants. If, in a survey

* Japan Advanced Institute of Science and Technology
† Japan Advanced Institute of Science and Technology

with 3, people judged that the chairman "intentionally" helped the environment, while he did *not* "intentionally" harmed the environment in the sense of 3, then that would show that the standard Knobe case was actually (at least mainly) an effect of the *linguistic* concept of intentional action peculiar to the *English* expression for intentionality, rather than a psychological effect, contrary to what Knobe (2016) suggested.

Unfortunately, or fortunately for Knobe, what we observed was exactly the same pattern observed in the standard Chairman case for English speakers (and speakers of other languages). Thus, such results provide crucial evidence that the role of the linguistic factor in the Knobe effect is limited. Even though there should be some such influence, that can be easily overwhelmed by the strong psychological effect. We should not overestimate the linguistic effect, at least when it comes to the Knobe effect. The moral asymmetry we find there is largely psychological.

We shall briefly discuss to what extent this conclusion can be generalized to other possible and actual cases. For each effect found in the fully contextualized use of the relevant terms, we can examine the robustness of the linguistic effect involved in it, and there are indeed clear cases in which the large effect observed in a questionnaire with a vignette may be wholly linguistic, with virtually no psychological effect involved. For example, Japanese has two distinct verbs for propositional knowledge, whose behavior can be very different in some contexts. Thus, the judgments about whether an agent knows something or not can differ radically in epistemologically interesting cases (Mizumoto 2018b). But if so, since they are judgments by the same people about the same cases, the effect is linguistic.

Thus, which effect, psychological or linguistic, is dominant in the data of a strong effect is just an empirical question, depending on the specific effect in question. Perhaps the psychological effect was so robust in the case of the Knobe effect because morality is evolutionarily more basic than other factors. But in other cases, the strength of the psychological effect in relation to the linguistic effect (of corresponding lexical items) within an overall effect, may differ from culture to culture, or language to language, each effect of which therefore still deserves a systematic investigation.

References

Knobe, J. (2016). Experimental philosophy is cognitive science. In Sytsma, J., & Buckwalter, W. (Eds.). (2016): 78-96.

Mizumoto, M. (2018a). A simple linguistic approach to the Knobe effect, or the Knobe effect without any vignette. *Philosophical Studies*, 1-18.

Mizumoto, M. (2018b). "Know" and Japanese Counterparts: "Shitte-iru" and "Wakatte-iru". In M. Mizumoto, S. Stich, and E. McCready (eds.), *Epistemology for the Rest of the World*. Oxford: OUP. 77-122.

Determining the boundaries of "Agent-hood": the role of Control-Asymmetry

Maria M. Piñango and Muye Zhang
Department of Linguistics
Yale University

The problem. Within linguistics, *agency* is observed through the semantic role of the *agent,* a role normally associated with conceptual functions such as willful undergoer of an action e.g. Ana in "Ana runs", or as actor in a causal event e.g., Ana in "Ana broke the vase". Indeed, the standard diagnostic for "agency" is the felicitous use of adverbial modifiers such as "voluntarily" or "on purpose", making salient the volitional engagement of the participant in the action. Whereas semantic roles are said to be licensed by specific predicates, the felicitous assignment of the role involves also the properties of the referent that bears the role. Consider the following sentences:

1. (a) Ana came in to the room and broke the vase (on purpose/voluntarily)
   (b) The ball went through the window and broke the vase (*on purpose/voluntarily)
   (c) The wind violently opened the door and broke the vase (*on purpose/voluntarily)

Whereas in all three sentences an individuated entity associated with the subject acts on the vase causing the vase to change state to 'broken', only in (1.a) can the semantic role of 'agent' be felicitously assigned. Inanimate entities "the ball" or natural forces "the wind" can participate in those events yet they fail the diagnostic test, suggesting instead that agent-hood assignment may not be categorical but a matter of degree (e.g., Grimshaw, 1990, Jackendoff 1990, Levin & Rappaport-Hovav, 1995). Consider also (2) below:

2. (a) The girl rode from Boston to New York (on purpose/voluntarily)
   (b) The bus rode from Boston to New York (*on purpose/voluntarily)

Here again only in (2.a) can agency be assigned to the NP subject even though in both cases the same change of location obtains. The difference is due to our expectation that an *agent* is a participant that is willfully performing the action in question, and not just undergoing the change of location. And this is independent of animacy. For example, in (1.a) above "Ana" would not be labelled an "agent" if she were to break the vase inadvertently. Predicates of possession offer yet another interaction with agent-hood. Consider:

3. (a) The girl owns/has a dog
   (b) The dog belongs to the girl

The possessors in both (3.a) are (3.b) fail the 'agent' diagnostic e.g., "?? The girl owns the dog on purpose" or "?The dog purposely belongs to the girl ". That is because predicates of possession are stative in nature and the "on purpose" adverbial takes events as its arguments. Yet for some possessor-possession relations, particularly those involving alienable possession with a human possessor, volition of the kind present in standard agents *is* observed. This is evidenced in the alienable vs. inalienable contrast: "The girl chose to have a car" vs. "??The girl chose to have a spleen". Even though in both cases possession is observed, in the second one the will of the possessor appears relevant to the truth of the prejacent [girl have car] thus indicating a kind of agency on the part of the subject NP referent.

So, on the one hand the role of *agent* appears intuitive and categorical and on the other its "boundaries" of assignment are elusive. This is seen in assignments that appear viable but are not warranted e.g., causatives with inanimate subjects, and in assignments that appear nonviable but are warranted e.g., possessors in alienable possession constructions. The overarching observation is that agent-hood assignment is not categorical, but resulting from the convergence of diverse factors some of them gradient. At the same time historically, agent-hood has played a key role in most theories of argument structure linking, where it has been used as the baseline against which all other semantic roles e.g., experiencer, source, patient, theme etc… are defined (e.g., Bresnan & Kanerva, 1989; Chomsky, 1981; Grimshaw, 1990; Jackendoff 1990, 1997). What is needed then is a *unified conceptual model from which we can predict the argument structure-syntax linking regularities observed without having to give up the intuitions underlying "agent-hood" that gave rise to it in the first place.* That is the aim of our proposal.

The proposal. We argue for a model of conceptual organization whereby our perception of agent-hood, and consequently the role that we assign to participants in an eventuality is an emergent effect resulting from the interaction of two factors: perceived or expected control-asymmetry and connectedness between participants in a situation. We call the space that the factors parametrize, the Multidimensional Space Model (MdS Model) (Piñango, 2019): Control-asymmetry and connectedness organize semantic memory, a long-term episodic memory space of storage, generalization and evaluation of situation-episodes: time/space-stamped percepts involving individuated entities.

Control asymmetry has its roots in asymmetric force dynamics. It refers to the degree to which one entity is endowed with the potential to decide on the fate of another. Given two entities in a situation, high control asymmetry signifies an expectation of large power asymmetry between two entities. Low control asymmetry signifies little or no perceived power differential between two entities (e.g., Klein & Perdue, 1992, Talmy, 2000, Piñango, 2019). From less to more asymmetry: two wheels of a car→a car and a ball→ a person and a car. **Connectedness** refers to the degree to which participants in a situation are functionally part of each other such that pulling them apart would risk their ability to function e.g., from less to more connectedness: a person and their incidental location→ a student and her school→ a person and her house→ a person and her hair → a person and her brain.

We argue for the validity of the model by showing how it solves the "many-meanings" problem associated with English *have* previously claimed to resist generalization (Zhang, 2021): whereas prototypically, *have* can be used for a possession interpretation, however, even within possession, further contextualization reveals both alienable (Ana has a car) and inalienable (Ana has a liver) possession readings one the one hand, and coincidental (The maple tree has a car under it) and non-coincidental location (The maple tree has a nest in it) on the other. The parametrized space allows an account whereby possession can be captured as involving location coincidental or not along various degrees of control asymmetry. In this way it unifies into one meaning space cognitive constructs previously thought to be categorically distinct: existence, location, alienable and inalienable possession. Finally, we present behavioral (questionnaire) and neurological (fMRI) evidence for the sensitivity of the brain to the changes in interpretation of English *have*, from possession to location readings, as a function of changes in the control asymmetry between participants; that is, within the same predication space.

On this view, whether a possessor in an alienable situation such as "Ana has a car" can be referred to as an *agent* of possession, is not relevant, what is relevant is that being the controller of the car gives Ana *agent*-potential. Within the traditional semantic role system, such potential which is part and parcel of our understanding of *agency* would remain unacknowledged along with our expectations of the possible "agencies" that, as a result, Ana can exert with respect to the car: sell it, destroy it, give it away, modify it somehow. And crucially, these are *relational* expectations, that modulate interpretation at the larger discourse context of the sentence. The sense of agency does not arise in low control-asymmetry cases such as "Ana has a spot" (low connectedness) (as in Ana occupies a spatial extent by virtue of her existence), or "Ana has a brain" (high connectedness) where there is no expectation that she is able to control either in any way.

This approach to agenthood naturally captures the composite and gradient-like nature of agency: The best "agents" are those that can exert high control asymmetry and low connectedness e.g., a human and an inanimate entity that is also functionally disconnected from it. It is this configuration that specific linguistic predicates leverage. As the relative control asymmetry relation of the participants change, the potential for agency changes accordingly. Therein lies the gradient nature of the agent-hood property.

We conclude that by understanding agenthood as the result of the interaction of control asymmetry and connectedness we are able to maintain the intuition of a prototypical agentive situation while allowing for the shared properties of other *agent*-y situations not as exceptions, but as natural products of the continuum of a parametrized meaning space.

Contact info: maria.pinango@yale.edu/ muye.zhang@yale.edu

Bresnan, J., & Kanerva, J. M. (1989). The Thematic Hierarchy and Locative Inversion in UG In E. Wehrli & T. Stowell, (Eds.), Syntax and Semantics 26. San Diego, CA: Academic Press. **Chomsky, N. (1981).** Lectures on government and binding. Dordrecht:Foris **Grimshaw, J. (1990)** Argument Structure. MIT Press. **Jackendoff, R. (1990)** Semantic Structures. MIT Press. **Jackendoff, R. (1997).**The architecture of the language faculty. Linguistic Inquiry Monograph. Cambridge:MIT Press **Klein, W. and Perdue, C. (1992).** Utterance Structure: Developing Grammars Again. Studies in Bilingualism, Vol. 5. John Benjamins. **Levin, B., & Rappaport-Hovav, M. (1995)** Unaccusativity. Cambridge, MA: MIT Press **Piñango, M.M (2019)** Concept Composition during Language Processing: a Model and Two Case Studies. The Routledge Handbook of Applied Chinese Linguistics, Huang, Jing-

Schmidt and Meisterernst (eds.) Chapter 39 pp 624-644. **Talmy, L (2000)** Toward a Cognitive Semantics. Cambridge, MA: MIT Press **Zhang, M. (2021)** Linguistic Variation and Cognitive Variability: the case of English *have*. Yale dissertation.