

Intentions and the Content of Scalar Implicatures

In this talk, I focus on a common equivocality in how the content of scalar implicatures is specified. On the one hand, scalar implicatures are often specified as the *contents* of the beliefs of a speaker: $\neg\psi$ (e.g. ‘John did not drink all the beers’). On the other hand, implicatures are also specified as the *beliefs* (or other intentional states) of a speaker herself: BELS($\neg\psi$) (e.g. ‘The speaker believes that John did not drink all the beers.’).

If we look at the literature, we can notice that many authors (see e.g. Carston 1998, Sauerland 2004) use the specifications interchangeably. In my talk, I argue that there is a substantial difference between the belief specification and the content specification. The main argument for taking the distinction seriously is that they have different consequences for how the hearer plans her future actions.

In particular, and in contrast to the standard Gricean approach, I argue that BELS($\neg\psi$) is the preferred option only in specific contexts in which the management of expectations about the future actions of the speaker is required for hearer’s planning of her own future actions. Deciding whether I should vote for the president in the upcoming elections or whether I want to go with someone on a family holiday are among the paradigmatic examples. However, situations in which BELS($\neg\psi$) is preferred are far from ubiquitous and the content specification $\neg\psi$ can provide sufficient information for hearer’s planning of her future actions in many other contexts, e.g. in one of encounters.

If the distinction between the specifications is substantial, then the question arises: What is the relation between them with respect to their derivability? In the second part of my talk, I discuss a serial derivation of the content specification from the belief specification as a preferred option from the Gricean perspective and I argue that this option is not available in the case of imperatives. As I argue, the commitment-based approach (Geurts 2019a, 2019b) can explain how the content specification is derived in the case of imperatives without relying on the belief specification and thus it has an advantage over the standard Gricean approach.

The main tenet of the commitment-based approach is that the primary aim of communication is to establish commitments and, by doing this, to help the speaker and the hearer to coordinate their actions better. According to Geurts (2019a), by making utterances, the speakers commit themselves to act on the truth of an expressed proposition. On the one hand, the

speakers commit themselves to avoid such actions as would make the proposition false. On the other hand, the speakers commit themselves to act proactively in such a way as will make the proposition true. As Geurts (2019a: 9–11) argues, the analysis of utterances in terms of commitments can be applied to imperatives as well. For example, if the speaker commands John to drink some of the beers by uttering ‘Drink some of the beers!’, then the speaker commits herself to act on the truth of ‘John will drink some of the beers’.

As I argue, the analysis of imperatives in terms of commitments allows us to derive the scalar implicature ‘Do not drink all the beers!’ without the need to ascribe implausible beliefs to the speaker in the intermediate step of deriving the belief specification.

References

- Carston, R. 1998. Informativeness, relevance and scalar implicature. In *Relevance Theory: Applications and Implications*, 179–238.
- Geurts, B. 2019a. Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics* 45: 1–30.
- Geurts, B. 2019b. What’s wrong with Gricean pragmatics. In *Proceedings of ExLing 2019*, 1–8.
- Sauerland, U. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27: 367–91.

Matej Drobňák Ph.D.

Department of Philosophy and Social Sciences
Philosophical Faculty
University of Hradec Králové

WHY CAN'T WE JUST BELIEVE WHAT WE WANT?

INFERENCE, METACOGNITION, AND EXECUTIVE CONTROL

Wade Munroe

University of Michigan

Department of Philosophy and the Weinberg Institute for Cognitive Science

Consider a puzzling feature about mental life. There are certain mental states that we can form as a direct result of our intending to do so. For instance, on a whim, we can imagine that we are 10 feet tall. However, occurrent belief isn't like this. We can't just occurrently believe we are 10 feet tall as a direct result of intending to do so. What explains the disparate levels of control we can exert over imagining a proposition as compared to occurrently believing it? In this paper, I argue that this difference in control is a function of metacognitive monitoring and control procedures. My view is attractive not only because it comports with cognitive psychological literature on metacognition, but it also provides an explanation that easily translates to other extant disparities in mental control; for instance, although we can, on a whim, imaginatively simulate the experience of a past event, we can't make it seem as if we are remembering the event. In the following, I begin with a discussion of memory, as much of the literature on metacognition focuses on memory. I then discuss occurrent belief and conscious deliberation, using work in metacognition to explain why we can't believe a proposition on a whim. In closing, I contrast my account with the two most prominent philosophical accounts of the disparate levels of control we can exert over imaginings and believings.

Before beginning our discussion of memory, let's first introduce a bit of terminology. Let's let an episodic simulation be a multimodal representations of an event from a perspective. Given our differential ability to episodically remember and imaginatively simulate an event through episodic simulation—and given the distinct functional roles these cognitive processes play—one may suspect that episodically remembering and imaginatively simulating must be relatively encapsulated to keep the two processes separate and easily distinguishable. However, episodic memories and imaginative simulations are largely subserved by the same neural activity and cognitive processes (Addis, 2018; Schacter, Addis, & Buckner, 2007). Deciding whether a given episodic simulation of an event, *E*, constitutes an episodic remembering or imaginative simulation of *E* is no trivial cognitive task, a fact that is further supported by the frequency with which we error in determining whether an (aspect of) an episodic simulation constitutes the memory of a previously experienced event or the imaginative simulation of a never before experience counterfactual (De Brigard, 2014).

In sum, episodic simulations do not come with some magical propositional tag appended to them that indicates their source, in other words, information about the etiology of the representation, e.g., whether the representation is the reinstatement of a previous experience (a memory) or the generation of a novel, counterfactual representation (an imagined event). According to the dominant source monitoring framework, various metacognitive monitoring procedures track myriad features (or cues) of episodic simulations and their process of generation in determining the representations' source. What matters for our concerns is that one of the features used by metacognitive monitoring procedures is processing fluency, or the efficiency with which a representation is tokened. As compared to an episodic memory, generating an imaginative simulation of a novel event requires extensive use of the frontoparietal control network that undergirds the executive functions of domain-general cognitive flexibility and the inhibition of automatically activated associations. So, say you attempt to episodically simulate some event that you have never experienced. Metacognitive monitoring processes will pick up on the increased working memory demands—relative to simulating a previously experienced event—required to generate the representation. This will result in metacognitive monitoring processes tagging the episodic simulation as imaginative.

Clearly, we don't have control over the amount of working memory resources required to generate a given episodic simulation (this will be a function of the associative strength between the various elements of the simulation). Thus, we won't be able to, at will, generate a simulation of some event that is tagged as an episodic memory by metacognitive monitoring and control processes. In sum, we can generate an imaginative simulation of an event at will; however, try as we might, we can't make the imaginative simulation seem like a memory just by intending to do so.

Traditionally, metacognitive research has focused on metamemory processes that monitor and control mnemonic processing. However, more recently, Ackerman and Thompson (e.g., 2017) have generated a model of meta-reasoning processes that monitor and control reasoning. Ackerman and Thompson argue that similar cues

used in metamemory are used in meta-reasoning to tag the propositional representations involved in reasoning. The key metareasoning tags that are important for our concerns are feelings of certainty (hereafter, FOC). As Ackerman and Thompson write, “[meta-reasoning] monitoring processes give rise to states of certainty and uncertainty” (Ackerman & Thompson, 2017, p. 613) throughout deliberation. As one reasons, meta-reasoning monitoring processes result in FOC tagged to mental states. In turn, FOC determine the function of the mental states. A propositional representation tagged with a sufficiently high FOC will function as an occurrent belief in virtue of terminating reasoning and driving further behavior, e.g., one using the representation to guide overt action.

It's well established that FOC, like metamemory tags, are a (partial) function of processing fluency (Unkelbach & Greifeneder, 2013). A propositional representation tokened in working memory, in the cognitive context of an act of problem solving, that is fluently processed will typically engender a strong FOC and will, thus, likely function as an occurrent belief. So, as a direct result of desiring to do so, we can imagine that some arbitrary proposition is the case; however, (as with metamemory processes) we don't have direct control over the FOC tagged to the propositional representation that determines whether the representation functions as an occurrent belief. Thus, the explanation for the discrepancy in control we exert over imagining a proposition as compared to occurrently believing it mirrors the explanation for the discrepancy in control we exert over imaginative simulations as compared to episodic memories.

In contrast to my position, the two most prominent philosophical accounts explain the disparate levels of control we exert over imagining a proposition as compared to occurrently believing it by appeal to, respectively, the (supposed) facts that (i), as Bernard Williams famously put it, “belief aims at truth” (1970, p. 136) and (ii) truth is a constitutive normative standard for belief. It's generally accepted that belief aims at truth, in the relevant sense, insofar as we possess a personal-level aim to believe a proposition only if it is true (Velleman, 2000). Thus, what (supposedly) explains the disparate levels of control we can exert over imagining a proposition as compared to occurrently believing it is our acceptance of a personal-level aim to believe a proposition only if it is true. Similarly, the putative fact that truth is a constitutive normative standard for belief (in that you ought to believe that p only if p is true) is supposed to explain the disparate levels of control in virtue of our acceptance and utilization of the truth norm in guiding our deliberative practices (Shah, 2003). Thus, both philosophical accounts appeal to our person-level acceptance of some aim or norm to explain the discrepancies in control that we exhibit over certain mental states. However, as others have noted (Sullivan-Bissett, 2017), both accounts fall victim to a similar concern: Our possession of an aim or acceptance of a norm doesn't necessitate that the aim or norm guide behavior in all cases, given we possess a strong desire to act against the aim or norm in circumscribed circumstances; however, we just can't seem to get ourselves to believe a proposition just by strongly desiring to do so. My appeal to metacognitive monitoring and control procedures, on the other hand, doesn't appeal to any personal-level aims or acceptances of norms on the part of believing agents.

Works Cited

- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607-617.
- Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, 48(2-3), 64-88.
- De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155-185.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, 8(9), 657-661.
- Shah, N. (2003). How Truth Governs Belief. *Philosophical Review*, 112(4), 447-482.
- Sullivan-Bissett, E. (2017). Aims and exclusivity. *European Journal of Philosophy*, 25(3), 721-731.
- Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11-32): Psychology Press.
- Velleman, D. (2000). On the aim of belief. In D. Velleman (Ed.), *The Possibility of Practical Reason* (pp. 244--281): Oxford University Press.
- Williams, B. (1970). Deciding to believe. In B. Williams (Ed.), *Problems of the Self* (pp. 136--151): Cambridge University Press.

THE ETHICAL FUNCTION OF INTENTIONAL ACTION

MIKAYLA KELLEY

Stanford University

On a prominent view in philosophy of action, intentional action admits of a naturalistic analysis. We can state what intentional action is without setting foot in the domain of practical normativity. On one way of precisifying this view, intentional action is behavior that is embedded within a particular sort of causal nexus, a nexus which is difficult to specify but which we could in principle specify in purely naturalistic terms. In this way, we might say that the concept of intentional action is an empirical rather than normative concept.

Granted, all would agree that intentional action is an empirical concept that is very relevant to practical normativity and, in particular, morality. Indeed, intentional action is the central kind of behavior for which we hold one another morally responsible. But the concept of intentional action is like the concepts of, e.g., happiness, pain, and hunger in the following way: it is very relevant to morality but is not itself intrinsically moral; the nature of intentional action—like the nature of happiness, pain, and hunger—might be just as fruitfully studied with the usual methods of, e.g., the biological sciences.

This naturalistic picture has been challenged in numerous ways,¹ and here I will offer another such challenge. I argue that intentional action is, in a sense, intrinsically ethical. As we will see, in contrast to other ways of challenging the naturalistic picture, the thesis I advance is consistent with much of the naturalistic picture and thus, I suggest, should be appealing to its proponents. The central claim that I defend—the *Intrinsicality Thesis*—is the following two-part thesis spelling out the precise sense in which intentional action is intrinsically ethical. Here and throughout I underline words when referring to the associated concept rather than the associated phenomenon.

(Intrinsicality Thesis) Intentional action is intrinsically ethical in the following sense:

- a. (*Functional Role Thesis*) intentional action plays a functional role in our ethical conceptual scheme²
- b. (*Extensionality Thesis*) the extension of intentional action is fixed, at least in part, by this ethical functional role

Two immediate clarifications. First, by our ethical conceptual scheme, I mean the complex of concepts that tend to guide our thought and practices surrounding normative assessment of practical life, where normative assessment is understood broadly to include everything from deontic assessment to holding responsible and practical life is a matter of what is, in some broad sense, done by individuals that can act. Second, the Functional Role Thesis does not only claim that intentional action is a *component* of this ethical conceptual structure—a claim that would hardly be surprising—but also that it *supports* the proper functioning of the structure by playing a particular function relative to it. In this way, intentional action is different from, say, happiness which is plausibly a component of our ethical conceptual scheme (in light of its centrality to human flourishing) but plausibly does not play a supportive role relative to it.

Here is a sketch of my argument for the Functional Role Thesis: we start with a description of a functional role that a concept might play in our ethical conceptual scheme—a role I call the *Prioritizing Role*. We note

¹As in, e.g., Chisholm 1966, Knobe 2003, Wilson 1989, Schapiro 2001, Tenenbaum 2007. For discussion, see Bishop 1989.

²I thus offer a practical explication of intentional action, much in the spirit of Craig's (1991) practical explication of knowledge and Queloz's (2022) practical explication of voluntary action.

why it would be useful to have a concept that plays the Prioritizing Role. We then infer from this usefulness that we in fact have a concept that plays this role. Next, we consider the features that a concept that plays the Prioritizing Role would have. Finally, we note that these features look very much like the most basic, agreed upon features of intentional action. We infer therefore that it is intentional action that plays the Prioritizing Role.

Here is a bit more about the Prioritizing Role. Things that act do a great many kinds of things. Actors do things on purpose, intentionally, voluntarily, freely, on accident, negligently, unknowingly, impulsively, automatically, habitually, uncontrolledly, and the list goes on. Doings of all kinds could, in principle and in the right context, be up for various forms of ethical evaluation. Doings of all kinds could be evaluated with either evaluative or deontic concepts—such as ‘good’, ‘bad’, ‘impermissible’, and ‘ought’—or they could be subject to praise, blame, and other mechanisms of holding responsible. Your accidentally insulting me might be deemed something you ought not have done, a dog doing a trick at the command of their owner might be deemed good, one’s having a seizure might be deemed bad, and I might hold you responsible for negligently but accidentally leaving the milk out to spoil. Given that there are so many things which could, in principle and in the right context, be up for some sort of ethical evaluation, it would be extremely useful to have a concept to focus our evaluative efforts—a concept which, when applied by oneself or others we are in dialogue with, makes clear that the doing in question is (*ceteris paribus*) “up first” for ethical evaluation broadly construed. In other words, this concept would play the functional role of flagging a special subset of all things done that are of *priority* for ethical evaluation broadly construed. This functional role that a concept might play is what I call the *Prioritizing Role*.³

As for the Extensionality Thesis, I argue for it from the Functional Role Thesis and the view that an action kind is specified by the kind of control that underlies it and a threshold of sufficient control of that kind.⁴ One reason to be skeptical of the Extensionality Thesis is that it seems to get the order of explanation backwards: arguably there is “out in the world” independent of our concepts a certain kind of thing—a certain joint in nature—and it is because of the nature of that thing that the concept which picks out that kind plays the functional roles that it does. The basic thought here is that this point about explanatory order is only partially right. Out in the world independent of our thinking about it, there are control kinds. We then have action concepts that cling onto some of these control kinds because having concepts that delimit those control kinds serve certain human purposes. But while there exist control kinds out in the world independent of us and the nature of these kinds explain the roles that our action concepts play, there do not exist thresholds of sufficient control out in the world independent of us. Instead, the threshold of sufficient control associated to a given action kind is determined by the purposes for which we have a concept which clings onto the control kind that underwrites that action kind.

Now recall the Functionality Thesis: one of the purposes of intentional action is to play the Prioritizing Role. Thus, the threshold of sufficient control associated with intentional action (the action kind) is influenced by intentional action playing the Prioritizing Role. It is in this way that the extension of intentional action is fixed, at least in part, by its ethical functional role. Thus, we arrive at the Extensionality Thesis.

Finally, we note that on the proposed way of seeing intentional action as intrinsically ethical, we do not posit any non-naturalistic entities like agent-causation or irreducible teleology; nor do we assume the existence of non-naturalistic normative facts about reasons or rationality. Instead, intentional action is intrinsically ethical in the sense that we have certain ethical practices up and running, and facts about how to streamline these practices—facts which are perfectly naturalistic—influence what counts as an intentional action.

³Note that to play the Prioritizing Role is not just to be a member of our ethical conceptual scheme but also to support its proper functioning; in particular a concept that plays the Prioritizing Role helps to direct and streamline the use of our ethical conceptual scheme.

⁴I defend this view of action kinds in work in progress (Kelley ms).

REFERENCES

- John Bishop. *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge University Press, 1989.
- Roderick Chisholm. Freedom and action. In Keith Lehrer, editor, *Freedom and Determinism*. Random House, 1966.
- Edward Craig. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Clarendon Press, 1991.
- Mikayla Kelley. A control theory of action. ms.
- Joshua Knobe. Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16(2):309–325, 2003.
- Matthieu Queloz. The essential superficiality of the voluntary and the moralization of psychology. *Philosophical Studies*, 179(5):1591–1620, 2022.
- Tamar Schapiro. Three conceptions of action in moral theory. *Noûs*, 35(1):93–117, 2001. doi: 10.1111/0029-4624.00289.
- Sergio Tenenbaum. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge University Press, 2007.
- George M. Wilson. *The Intentionality of Human Action*. Stanford University Press, 1989.

INTENTION AND ABILITY: A CHALLENGE FOR MENTALISM

Aznavur Dustmamatov (*Independent*)

Introduction. I argue that the ability to act is a constituent part of the state of intention, and that abilities are non-mental things, since (roughly) one cannot acquire an ability by representing oneself as having it. As a result, I contend that intentions are not reducible to mental states such as beliefs.

Intention and ability. Annette Baier claimed that one cannot intend things that one cannot do: “If I cannot play the harpsichord I cannot intend to play it; at most I can intend to learn to play it.” (Baier 1970) I argue that, properly developed, Baier’s claim is true, but there are three qualifications that must be mentioned:

Attempted action. That I must be ‘able to Φ ’ does not imply that I must be able to successfully complete Φ -ing. Instead, I must be only able to successfully *attempt* Φ -ing. Successful Φ -ing need not even be possible: I can intend to Φ even if it is *impossible* to (successfully) Φ , provided that I can ψ and ψ counts as trying-to- Φ (Ludwig 1992; 1995). Of course, what counts as ‘attempt’ cannot be arbitrary: e.g. I must be at least a decent piano player in order to attempt playing the Italian Concerto.

Minimal ability. Some may object to the requirement that one have the ability to try, since I could ψ , even if I do not have the *ability* to ψ . I might not have the ability to hit 10 in darts, yet on occasion, out of sheer luck, I hit 10 with a dart. Even in this case, I must have the ‘minimal ability’ to hit 10 (Costa 1986), which is to say I must have the ability to swing my hand, to hold the darts in my fingers, and so on; and my doing those things with an intention to hit 10 increases my chance of hitting 10.

Agential control. It is also required that I have ‘agential control’ over the exercise of this minimal ability, i.e. that it is not a mere disposition or coincidence. After all, the intention can increase the chance of a certain outcome, if, say, God arranged the world in such a way as to conform with my intentions; still, this wouldn’t count as intentional action, since it does not issue from the *exercise* of *my* ability.

In summary: I can intend to Φ , iff I have the (minimal) ability to ψ , where ψ counts as trying-to- Φ , and I have agential control over this ability.

Necessity and parthood. I argue that the minimal ability to attempt (in the above sense) is not only a necessary condition for intention, but also a constituent part of the state of intention. We must understand ‘intention’ in the right sense, however, as the term can mean two things: *what* is intended, and the (token) *act* of intending it. The two are distinct, as there can be an intention (such as ‘to depose the king’) that either nobody happens to be intending, or many are intending at the same time.

Although the ability to act is not a constituent part of intention in the first sense as it usually does not figure in the descriptions of what is intended, it is a necessary constituent of any token act of intending. Constituent parts of x are typically such that without these parts the whole does not count as x : a ‘table’ without any legs or support, a mere tabletop laying on the floor, would not count as a table at all. By this criterion, the minimal ability to attempt counts as a constituent part of the token act of intending (or ‘having an intention’), since without it, merely believing in a future action does not amount to a genuine intention, as the analysis in the previous section has established.

Ability as non-mental. I argue that abilities are not reducible to mental states. What makes some states mental rather than, say, bodily is that the content of representation determines the state as what it is: one desires x by representing x as desirable, one believes that p by representing the world as p , and so on. Abilities are different in that representing Φ as doable does not of itself make Φ doable, and representing oneself as being able to Φ does not entail one’s ability to Φ : one may think oneself able, yet still be unable.

The mentalist could argue that, although this is true for most abilities, it is not true of mental abilities, which are the abilities to take mental actions, such as counting to ten in one’s head. What

makes such actions mental is a muddled matter, as philosophers resort to providing examples of typical mental actions, instead of defining what makes them mental (Peacocke 2021). Even if we concede that there are genuinely mental actions, the ability to perform those actions need not be a mental state (or its part): after all, even if I represent myself as being able to count to ten in my head in German, I may in fact be unable to do so.

Possible objections and replies. The mentalist may concede that ability is a non-mental thing, or at least something that is not fully mental, but still argue that this non-mental thing is a constituent of a state that is mental as a whole, provided that the ability is contained in the *content* of this mental state. This is based on the externalist thesis that a non-mental thing can become part of a mental state if it figures in its contents, such as when XYZ (a physical substance) becomes part of a mental state in the sense that there is, somewhere, a mental state whose content is XYZ.

Since intention requires that I am able to Φ as a matter of fact (i.e. in *this* world), the suggestion would work only if the contents of a mental state were *facts*, rather than propositions. However, this would imply that intention requires *knowledge* of my ability to Φ instead of a mere belief in it. But, although knowledge is argued to be a mental state by some (Williamson 2000; Nagel 2013), this view remains controversial among philosophers and has bizarre implications that go beyond anything that other kinds of externalism entail (Smith 2017).

Even if we concede that knowledge is a mental state, or that the (non-mental) ability can be one part of a mental state such as belief, there is yet another problem for mentalism to reckon with. The knowledge of, or even belief in, my ability to Φ does not seem necessary for intentional action (e.g. Davidson 1971; 1978). One family of cases concerns the disbelief in one's ability: for example, one may think one is unable to Φ , as in the situation of paralysis (Setiya 2008), yet still perform the action (against all expectations). The other family of cases concerns spontaneous or lightning-fast actions that are not unaccompanied by any particular belief about one's ability: for example, an athlete who performs a surprising feat by reacting in a split-second; or a hero bystander who, without further thought, jumps on the tracks to save another person.

Select bibliography. Baier, A. "Act and Intent." *Journal of Philosophy* 67 (1970): 648-58. Costa, M. "Acting Intentionally and Minimal Abilities." *Analysis* 46(3) (1986): 144-147. Ludwig, K. "Impossible Doings." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 65(3) (1992): 257-281. Ludwig, K. "Trying The Impossible: Reply To Adams." *Journal Of Philosophical Research* 20 (1995). Nagel, J. "Knowledge as a Mental State." *Oxford Studies in Epistemology* 4 (2013): 275-310. Peacocke, A. "Mental Action." *Philosophy Compass* 16(6) (2021): e12741. Setiya, K. "Practical Knowledge." *Ethics* 118 (3) (2008): 388-409. Smith, M. "The Cost of Treating Knowledge as a Mental State." In A. Carter, E. Gordon & B. Jarvis (eds.), *Knowledge First, Approaches to Epistemology and Mind*. Oxford University Press (2017): 95-112. Williamson, T. *Knowledge and its Limits*. Oxford University Press (2000).

Trying without fail

Ben Holguín* and Harvey Lederman†

This paper examines properties of trying. The first half of the paper discusses some questions about the relationship between trying on the one hand, and believing, desiring, and intending on the other. We extend important work by [Hornsby \(1995\)](#) and [Ludwig \(1992\)](#) to offer new arguments for the claims that a person can try to do something even if they believe it is impossible that they will succeed, that a person can try to do something even if they do not want to succeed, and that a person can try to do something even if they do not intend to succeed. The second half of the paper turns to the relationship between trying to do something and trying to try to do it. We argue that a person tries to do something if and only if they try to try to do it, and then explore some consequences of this claim, both for the logical structure of intentional action and the question of whether basic actions are tryings.

The basic arguments of the first half of the paper concern cases like the following. A person stands before a wall, which they are certain they cannot push over, and which they do not want to fall over, but tries to push it over, in order to prove that they cannot. We argue, with [McCormick & Thalberg \(1967, p. 45\)](#), [Harman \(1986, p. 370\)](#), [Hornsby \(1995\)](#), that this person can indeed try to push the wall over. A common complaint about this kind of case (see e.g. [O’Shaughnessy \(1973, 1980\)](#), [Jones \(1983\)](#), [McCann \(1986\)](#), and [Adams \(1995, 2007\)](#)) is that although it is natural to say that this person is trying to push the wall over, this is not because the claim is literally true; it is rather because it is true that the person is pretending to try or acting as if they are trying. We offer several new arguments against this diagnosis.

If these arguments are successful, the case provides an example of someone who tries to do something while being certain they will not succeed, while not wanting to succeed, and not intending to succeed. We also provide a new case in which a person tries to do something which they do not want to do, even when they do believe it is possible they will succeed. Finally, we leverage cases related to “Butler’s puzzle” ([Butler \(1978\)](#), cf. [Kraemer \(1978\)](#), [Ross \(1978\)](#), [Nadelhoffer \(2004\)](#)) in which a person tries to do something which they do not intend to do, even when they both believe it is possible they will succeed, and want to succeed.

The second half of the paper argues that a person tries to do something if and only if they try to try to do it. Building on our arguments for this claim, we argue also that a person tries to do something if and only if they intentionally try to do it. We develop three consequences of this latter claim. First, we

*Johns Hopkins University, Philosophy, bholgui1@jh.edu.

†University of Texas at Austin, Philosophy, harvey.lederman@austin.utexas.edu.

argue that if a person tries to do something, they intentionally...intentionally try to do it, suggesting a disanalogy between intentional action (which can non-trivially iterate indefinitely) and knowledge (which has been argued by Williamson (2000) not to). Second, we connect this property to the notion of basic action, and bolster others' arguments for the claim that basic actions are tryings. Third, we consider the question of what "options" are in normative decision theory and argue that, if it is true that, if one tries to do something, one does it, then tryings are natural candidates to be options in this sense.

References

- Adams, Frederick. 1995. Trying: You've Got to Believe. *Journal of Philosophical Research*, **20**, 549–561.
- Adams, Frederick. 2007. Trying with the Hope. In: Timmons, Mark, Greco, John, & Mele, Alfred R. (eds), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford University Press.
- Butler, Ronald J. 1978. Report on Analysis "Problem" No. 16. *Analysis*, **38**(3), 113–114.
- Harman, Gilbert. 1986. Willing and Intending. *Pages 363–380 of: Grandy, Richard, & Warner, Richard (eds), Philosophical grounds of rationality: Intentions, categories*. Clarendon Press Oxford.
- Hornsby, Jennifer. 1995. Reasons for Trying. *Journal of Philosophical Research*, **20**, 525–539.
- Jones, O. R. 1983. Trying. *Mind*, **92**(367), 368–385.
- Kraemer, Eric Russert. 1978. Intentional Action, Chance and Control: [Analysis "Problem" No. 16]. *Analysis*, **38**(3), 116–117.
- Ludwig, Kirk. 1992. Impossible Doings. *Philosophical Studies*, **65**(3), 257–281.
- McCann, Hugh J. 1986. Rationality and the Range of Intention. *Midwest Studies in Philosophy*, **10**(1), 191–211.
- McCormick, Suzanne, & Thalberg, Irving. 1967. Trying. *Dialogue*, **6**(1), 29–46.
- Nadelhoffer, T. 2004. The Butler Problem Revisited. *Analysis*, **64**(3), 277–284.
- O'Shaughnessy, Brian. 1973. Trying (As the Mental "Pineal Gland"). *Journal of Philosophy*, **70**(13), 365–386.
- O'Shaughnessy, Brian. 1980. *The Will: A Dual Aspect Theory*. Cambridge University Press.
- Ross, David. 1978. He Loads the Gun, Not the Dice: [Analysis "Problem" No. 16]. *Analysis*, **38**(3), 114–115.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.

Agent concepts diverge from agentive syntax: evidence from an event categorization study

Lilia Rissman¹, Sebastian Sauppe^{2,3}, Arrate Isasi-Isasmendi^{2,3}, Susan Goldin-Meadow^{4,5}, Balthasar Bickel^{2,3}

1. Department of Psychology, University of Wisconsin – Madison
2. Department of Comparative Language Science, University of Zurich
3. Center for the Interdisciplinary Study of Language Evolution, University of Zurich
4. Department of Psychology, University of Chicago
5. Department of Comparative Human Development, University of Chicago

Introduction Event roles such as Agent and Patient have been argued to be cross-linguistically universal and crucial for language evolution [1-3]. One challenge to this universal view is that Agent-marking syntactic structures in different languages express different semantic categories [4]. For example, intransitive (one-participant) verbs (e.g., *jump*, *arrive*, *die*) range on a semantic continuum from more activity-oriented (e.g., *jump*) to more state-oriented (e.g., *die*). In English, the arguments of activity-oriented verbs and state-oriented verbs are expressed in the same way (all are marked by nominative case). In Basque, by contrast, more activity-oriented verbs mark their arguments with what is known as ergative case, while the arguments of more state-oriented verbs are nominative-marked. We investigate whether these different syntactic systems correspond to English and Basque speakers conceptualizing Agency in different ways. Specifically, we test two ways in which Agent roles might differ. First, English and Basque speakers might represent Agent in terms of different prototypes. In linguistic theory, event roles are often analyzed in terms of proto-Properties: for example, being intentional and playing a causative role are properties of proto-Agents whereas being affected is a property of proto-Patients [5]. The proto-Properties that constitute Agency may differ for English and Basque speakers. Second, English and Basque speakers might diverge in how they conceptualize the single participant in an intransitive event (e.g., one who jumps, one who arrives) with respect to the Agent category. Consistent with how arguments of intransitive verbs are marked in these two languages, English speakers might represent an individual who *arrives* as more Agentive than Basque speakers do. We tested these hypotheses using an event categorization task in which participants learned to sort pictures of transitive (two-participant) events into Agent and Patient piles, building on Rissman and Lupyan [6]. At test, we asked participants to generalize these categories to transitive events with more or less prototypical Agents and Patients, testing our first question, and to generalize these categories to intransitive events, testing our second question.

Method We recruited 108 English speakers and 109 Basque speakers who completed the study online. In the training phase of the experiment, participants saw 28 images of one figure acting on another. Either the Agent or the Patient was shaded red (see Figure 1). Participants learned to group the pictures into “Agent” and “Patient” categories (labelled Category “A” or “B”), receiving accuracy feedback on every trial. Participants then completed a test phase where they viewed new images and decided whether the scenes belonged to Category “A” or “B”. This test phase included both transitive and intransitive scenes. The *transitive scenes* featured more or less prototypical Agents and Patients (e.g., the roles in Figure 1 being more prototypical; the roles in a scene of one person whispering to another being less prototypical). We used the prototypicality norms in Rissman and Lupyan [6], who normed the transitive scenes for six of Dowty’s proto-Properties: intentionality, causation, movement, change of state, affectedness, and being stationary. The *intransitive scenes* featured both

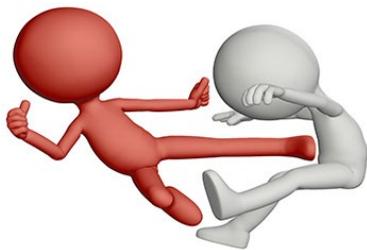


Figure 1. A sample training picture

activity-oriented events (e.g., jumping, running) and state-oriented events (e.g., someone grabbing their stomach as if sick); see examples in Figure 2. Across all participants, we tested 48 transitive scenes and 48 intransitive scenes. Each participant viewed 48 transitive trials (half with a red Agent and half with a red Patient) randomly interspersed with 24 intransitive trials (showing a single, red-shaded individual). No feedback was provided on the test trials.

In a separate online task, 20 English and 10 Basque speakers provided written descriptions of the 48 intransitive scenes. We coded each description as to whether activity or state-oriented verbs were used. For English, we annotated verbs as taking agent-like or patient-like subject arguments (technically known as “unergative” and “unaccusative” verbs, respectively) using the diagnostics from Momma, Slevc [7]. For Basque, we annotated whether the verbs assigned ergative or nominative case to their argument. For each intransitive scene in each language, we calculated an “agentivity score” – a weighted average of how often activity-oriented verbs were produced for that scene.

Results & Discussion Test accuracy for transitive scenes was high: English and Basque speakers correctly categorized the pictures into Agent and Patient categories on 90% of trials ($CI_{95} = [89\%, 91\%]$). Accuracy for individual transitive scenes was almost perfectly correlated across English and Basque speakers: $r(46) = .96$, $p < .001$. Accordingly, the same proto-Properties predicted generalization accuracy in the two languages. Participants were more accurate when the Agent was more intentional (English: $b = .45$, $SE = .14$, $p < .001$; Basque: $b = .44$, $SE = .12$, $p < .001$) and when the Agent caused the event (English: $b = .25$, $SE = .14$, $p = .07$; Basque: $b = .26$, $SE = .12$, $p < .05$). These results suggest that English and Basque speakers represent transitive event roles in highly similar ways.

Does this similarity extend to intransitive scenes, for which the two languages use opposing grammatical systems? On average, the Basque and English speakers classified the intransitive pictures into the

Agent category on 71% of trials ($CI_{95} = [69\%, 73\%]$). Crucially, rates of classifying individual scenes in the Agent category were strongly aligned across the two languages: $r(46) = .83$, $p < .001$; see Figure 2. The agentivity score of each picture in each language was positively correlated with how likely speakers were to classify that picture as an Agent (English: $r(46) = .44$, $p < .01$; Basque: $r(46) = .64$, $p < .01$). As these r -values show, however, the relationship between ergative case marking and Agent sorting ($r = .64$) was weaker than the relationship between English and Basque Agent sorting ($r = .83$). In addition, differences in agentivity scores between English and Basque did not predict differences in Agent sorting between the two languages ($r(46) = -.25$, $p > .1$). As with the transitive scenes, speakers of these two

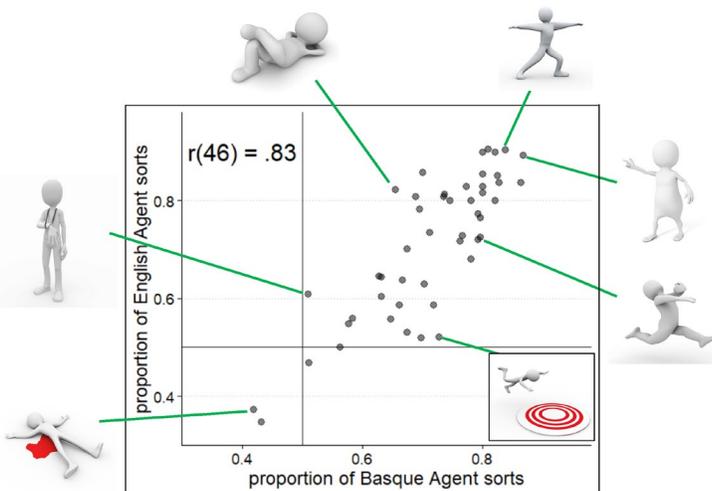


Figure 2. Rates of classifying individual intransitive scenes into the Agent category for Basque vs. English speakers

languages classified intransitive scenes in highly similar ways.

The similarity that we observe across English and Basque suggests that despite their radically different linguistic treatment of Agent-like and Patient-like arguments, they share semantic role representations that are not modulated by syntax. This finding supports theories in which event roles are cross-linguistically universal, despite different ways of encoding these roles in grammatical structures.

References

1. Strickland, B., *Language reflects “core” cognition: A New theory about the origin of cross-linguistic regularities*. *Cognitive Science*, 2017. **41**: p. 70-101.
2. Zuberbühler, K. and B. Bickel, *Transition to language: From agent perception to event representation*. *WIREs Cognitive Science*, 2022. **13**(6): p. e1594.
3. Rissman, L. and A. Majid, *Thematic roles: Core knowledge or linguistic construct?* *Psychonomic Bulletin & Review*, 2019. **26**(6): p. 1850-1869.

4. Comrie, B., *Language universals and linguistic typology: Syntax and morphology*. 1989, Chicago, IL: University of Chicago Press.
5. Dowty, D., *Thematic proto-roles and argument selection*. *Language*, 1991. **67**(3): p. 547-619.
6. Rissman, L. and G. Lupyan, *A dissociation between conceptual prominence and explicit category learning: Evidence from agent and patient event roles*. *Journal of Experimental Psychology: General*, 2022. **151**(7): p. 1707.
7. Momma, S., L.R. Slevc, and C. Phillips, *Unaccusativity in sentence production*. *Linguistic Inquiry*, 2018. **49**(1): p. 181-194.

German Rationale Clauses Track Intentional Attitudes – AIL 3

Felix Frühauf (University of Hannover/University of Konstanz)

Intro. Rationale Clauses (RatCs), are used to express someone’s rationale or the intention with which an event was brought about. In German, they are expressed in either finite (introduced by *damit*) or non-finite form (introduced by *um*). If the matrix clause describes an event that cannot be brought about by intentional action (or by natural design), the Rationale Clause is heavily degraded. A felicitous interpretation of (1b) would require Susi to have rigged the lottery in her favor. On the basis of facts like these, Rationale Clauses are often mentioned as tracking the RESP(onsibility)-relation in the sense of Farkas (1988).

- (1) a. *Susi hat sich einen Glücksbringer gekauft, um im Lotto zu gewinnen.*
Susi has herself a luck.bringer bought UM in.the lottery to win
‘Susi bought a talisman in order to win the lottery.’
- b. #*Susi hat im Lotto gewonnen, um ihre Schulden zurückzahlen zu können.*
Susi has in.the lottery won UM her debt pay.back to be.able
int. ‘Susi won the lottery to be able to pay back her debt.’

Attitude data. This generalization falls short of describing the whole range of observed data. Attitude verbs like *hope*, *want*, *wish*, or *dream* (as an ‘intentional’ priority attitude) admit of modification by Rationale Clauses, while they don’t necessarily require the attitude holder to be in the RESP-relation to their holding the attitude.

- (2) a. *Susi hofft, dass die Sonne scheint, damit sie draußen sitzen kann.*
Susi hopes that the sun shines DAMIT she outside sit can
‘Susi hopes that the sun will shine in order to be able to sit outside.’
- b. *Susi will, dass die Sonne scheint, um draußen sitzen zu können.*
Susi wants that the sun shines UM outside sit to can
‘Susi wants the sun to shine in order to be able to sit outside.’

I argue that the Rationale Clause does not track RESP-hood, but the availability of an ‘intentional attitude’, compatible with a variety of future directed preferences. Seen this way, the paradigmatic agentive action case (1a) is an outlier insofar as it doesn’t wear its attitude on its sleeves. But in line with recent work by Alonso-Ovalle and Menéndez-Benito (2018) and Alonso-Ovalle, Menéndez-Benito, and Rubinstein (2022), I assume that intentional action projects a modal domain similar to *want* or *hope*.

Interpretation. Following a common assumption, I assume that a Rationale Clause sequence ‘*p* UM/DAMIT *q*’ encodes the attitude holder’s belief that *p* enables the realisation of favorable *q* (see, e.g. Balkanski (1992)). This is easily illustrated for intentional action (3a). But a parallel treatment of the attitude cases in (2), leads us into trouble (3b). After all, it is not the hoping attitude that is believed to act as an enabler, but the realization of said hope. The correct paraphrase should thus be (3c).

- (3) a. (1a) \approx Susi bought a talisman and she believed that buying a talisman would make it more probable to win the lottery.
- b. (2a) $\not\approx$ Susi hopes that the sun will shine and she believes that hoping that the sun will shine would make it more probable to be able to sit outside.
- c. (2a) \approx Susi hopes that the sun will shine and she believes that the sun shining would make it more probable to be able to sit outside.

Generalization: The attitude holder of the matrix event believes that the best continuations of the attitude-content-becoming-true worlds are q -worlds. (4) presents a shot at formalizing this intuition, assuming with Hacquard (2006) that modal domains are projected from events. For *hope*, these are the worlds compatible with the current hoping event, and for agentive action, these are the worlds compatible with the intentions involved in bringing about the event. The Rationale Clause encodes that the attitude holder believes that if the world is such that it conforms with his intentional attitude, it is such that q will be enabled.

- (4) a. $\llbracket x \text{ buys a talisman} \rrbracket = \lambda e. \text{BUY}(x, t, e)$
 b. $\llbracket x \text{ hopes } p \rrbracket = \lambda e. \text{HOPE}(x, e) \wedge \forall w' \in \text{BEST}_{(f,g,e)}[p(w')]$
 c. $\llbracket \text{UM } q \rrbracket = \lambda P. \lambda e. P(e) \wedge \forall w'' \in \text{DOX}(x)[w'' \in \text{BEST}_{(f,g,e)} \rightarrow \diamond q(w'')]$
 d. $\llbracket x \text{ hopes } p \text{ UM } q \rrbracket =$
 $\lambda e. \text{HOPE}(x, e) \wedge \forall w' \in \text{BEST}_{(f,g,e)}[p(w')]$
 $\wedge \forall w'' \in \text{DOX}(x)[w'' \in \text{BEST}_{(f,g,e)} \rightarrow \diamond q(w'')]$
 e. $\llbracket x \text{ buys a talisman UM } q \rrbracket =$
 $\lambda e. \text{BUY}(x, t, e) \wedge \forall w'' \in \text{DOX}(x)[w'' \in \text{BEST}_{(f,g,e)} \rightarrow \diamond q(w'')]$

Two readings for attitudes. I have argued that the Rationale Clause can either directly access an attitude event or that it can recruit an attitude from an intentionally brought about event. In the special case, in which an attitude event is brought about intentionally, we expect two possible interpretations. And this is what we find. (5) illustrates the two readings using two naturally occurring examples with *träumen* ‘dream’.

- (5) a. *Um über bessere Trainingsbedingungen zu verfügen, träumt man von einem neuen Hallenbad.*
 UM over better training conditions to dispose, dreams one from a new hall.pool
 ‘In order to have better training conditions, one is dreaming of a new swimming pool.’
 b. *Um die Wolken aus dem Gemüt zu vertreiben, träumt man vom Sommer.*
 UM the clouds out the mind to expell dreams one from.the summer
 ‘In order to drive the clouds out of one’s mind, one dreams of summer.’

(5a) illustrates the reading we have already seen: The attitude holder believes that if their dreams of a new swimming pool were to be realized, they would have better training conditions. (5b) illustrates the ‘intentional’ reading: Here, having the attitude (dreaming of summer) is believed to help reaching the goal of lifting the spirits. *Hope* can be coerced in the same way: one could hope for more sunshine in summer in order to survive the winter. Note that doxastic attitudes like *believe* can only receive the latter reading, and only if the context allows it.

- (6) # *Susi glaubt, dass die Sonne scheint, damit sie draußen sitzen kann.*
 Susi believes that the sun shines DAMIT she outside sit can
 int. ‘Susi believes that the sun will shine in order to be able to sit outside.’

In order to receive any reasonable interpretation, (6) has to imply that Susi intentionally formed a belief. The Rationale Clause then reports on Susi’s (irrational) belief that being able to sit outside is a possible outcome of her adopting the belief that the sun is shining. This contrasts with the *hope* case in (2a), in which her being able to sit outside is understood as a putative consequence of her hope being realized, not of acquiring this hope.

To conclude, Rationale Clauses (in German) can modify both intentional action and preferential attitude events. Thus, they are not indicative of RESP-hood, but intentionality more general. I have formulated an analysis making use of an enabling paraphrase for Rationale Clauses.

- Alonso-Ovalle, L. and P. Menéndez-Benito (2018). “Projecting possibilities in the nominal domain: Spanish *uno cualquiera*”. In: *Journal of Semantics* 35.1, pp. 1–41.
- Alonso-Ovalle, L., P. Menéndez-Benito, and A. Rubinstein (2022). “Sub-eventive Modal Projection: The Case of Spanish Subjunctive Relative Clauses”. In: *Talk at NELS* 52.
- Balkanski, C. T. (1992). “Actions, beliefs and intentions in rationale clauses and means clauses”. In: *AAAI*, pp. 296–301.
- Farkas, D. (1988). “On obligatory control”. In: *Linguistics and Philosophy* 11.1, pp. 27–58.
- Hacquard, V. (2006). “Aspects of modality”. PhD thesis. Massachusetts Institute of Technology.

Making others move against their will: On causatives of motion verbs and other unergatives

Eva Neu, University of Massachusetts Amherst

Introduction. It is commonly assumed that direct causatives can only be formed from unaccusatives. Direct causativization involves adding to the description of the verbal event a new agent, the causer. Since unergative verbs already come with an agent θ -role, causativizing an unergative should thus result in an event description with two agents, a derivation expected to crash both semantically and syntactically. In fact, however, direct causatives of unergatives are robustly attested in several languages (Hindi-Urdu, Turkish, Sason Arabic, English). I show that these causatives are ordinary transitives, with the previous agent demoted to the internal argument position associated with a canonical patient-like interpretation. My research thus demonstrates the effect of relative degrees of agentivity on argument structure: prototypically agentive activities can be construed as non-agentive if agentive properties such as volition and causal power are instead ascribed to another event participant, the causer.

Empirical contribution. Direct causatives of verbs which normally have an unergative use are attested in languages such as Hindi-Urdu (1), formed with the affix *-aa* also regularly found on direct causatives of unaccusatives:

- (1) a. Rohan **naach** rahaa hai.
Rohan.M **dance** PROG.MSG be.PRS.3MSG
'Rohan is dancing.'
- b. Shama Rohan-ko **nach-aa** rahii hai.
Shama.F Rohan-ACC **dance-AA** PROG.F be.PRS.3MSG
'Shama is making Rohan dance/twirling him around (the dance floor).' (Bhatt and Embick 2017:124)

The vast majority of Hindi-Urdu unergatives causativizes. Their status as unergatives is confirmed by their ability to form impersonal passives and their inability to appear in reduced relatives (Bhatt and Embick 2017). There is equally clear evidence for the direct and monoclausal status of such causatives: they are interpreted as monoeventive, involving spatio-temporal contiguity, they do not – unlike indirect causatives formed with the affix *-vaa* – permit the addition of intermediate agents, and the causee cannot be modified by subject-oriented adverbs.

I argue that direct causatives of unergatives are ordinary transitives. Causativized unergatives, unlike their intransitive counterparts, can appear in reduced relatives (2), indicating that the causee is merged as an internal argument:

- (2) a. ***daur**-aa arkaa
run-PFV.MSG boy
*'the run boy'
- b. [Ravi-dwaaraa **daur-aa**-yaa gayaa] larkaa
Ravi-by **run-AA**-PFV PASS.PFV boy
'the boy chased by Ravi' (Bhatt and Embick 2017:124f.)

Moreover, transitives as well as unergatives with path arguments or cognate objects are unable to causativize (3):

- (3) a. Rohan tango **naach** rahaa hai.
Rohan.M tango **dance** PROG.MSG be.PRS.3MSG
'Rohan is dancing the tango.'
- b. *Shama Rohan-ko tango **nach-aa** rahii hai.
Shama.F Rohan-ACC tango **dance-AA** PROG.F be.PRS.3MSG
Intended: 'Shama is making Rohan dance the tango.'

This is as expected if the causee would have to be merged in the position that is already filled by the direct object.

Beyond Hindi-Urdu, a subset of verbs in Turkish and Sason Arabic that pass unergativity diagnostics equally form direct causatives, and I provide syntactic evidence for their status as ordinary transitives as well. Moreover, it is well-known that English allows unergative motion verbs to form transitives under certain circumstances (4):

- (4) a. Shama danced Rohan *(around the ballroom).

Biggs (2019) establishes that examples like (4) are syntactically transitives with a PP-adjunct and not small clauses, by demonstrating that the verb and the object form a constituent at the exclusion of the PP (5):

- (5) a. Mary waltzed John around the ballroom, and Sarah did so around the garden.

b. Waltz John though Mary did around the garden...

(Biggs 2019:7)

It remains an open question why these constructions require PP-adjuncts – or, as Biggs shows, other modifiers, modals or negation – to be felicitous. Telicity of the verb phrase has been shown to not be the decisive factor (Folli and Harley 2006). Despite this restriction, and the fact that English limits causativization of unergatives to motion verbs, I will treat direct causatives of unergatives in Hindi-Urdu, Turkish, Sason Arabic and English as belonging to the same natural class, all forming ordinary transitives with the causee as an internal argument.

Theoretical consequences. The present analysis seems at odds with standard views on argument structure: the internal argument position which the causee would occupy is associated with a patient θ -role, but an event participant performing a dance – or another activity described by a normally unergative verb in a given language – is commonly assumed to receive an agent θ -role. In other words, while in the intransitive it is the external argument performing the activity denoted by the verb (e.g., dancing), in the transitive it would be the internal argument, meaning that a certain kind of event participant seems to not be mapped consistently to the same syntactic position.

I propose to regard direct causatives of unergatives as instances of variable unaccusativity: in causative contexts, the unergative verb is coerced into an unaccusative behavior, such that the previous external argument comes to be realized as an internal one instead. Cross-linguistically, many verbs can vary between an unaccusative and an unergative use, often associated with a reduced or enhanced agentivity of the argument (Sorace 2000). E.g., variable case marking with some verbs in Tsova-Tush shows that an event participant performing one and the same activity can be mapped either to the external or the internal position depending on its degree of intentionality (6):

- | | | | | | | |
|-----|----|------|----------------------------|----|----|-----------------------------|
| (6) | a. | (as) | vuiž-n-as. | b. | so | vož-en-sO. |
| | | | 1SG.ERG fell.AOR-1SG.ERG | | | 1SG.NOM fell.AOR-1SG.NOM |
| | | | ‘I fell down, on purpose.’ | | | ‘I fell down, by accident.’ |

(Holisky 1987:105)

Crucially, the causee in direct causatives of unergatives obligatorily receives a deagentivized interpretation in all languages under discussion, being depicted as not being in control of the event or even performing the activity against their will. Hence, the external argument of the intransitive performs the activity agentively while the internal argument of the transitive does so in a patient-like manner, fully in line with the alternation seen in (6).

While it is known that variable unaccusativity is sensitive to agentivity, the present research thus demonstrates the effect of *relative* degrees of agentivity in particular. In order to realize a canonically external instead as an internal argument in languages like Hindi-Urdu, it does not suffice to reduce its agentivity; otherwise, (2a) should be felicitous under a deagentivized reading. Rather, what is needed to dethrone the agent is the presence of another agent, the causer. In causatives of unergatives, the latter typically needs to be animate, as seen for English in *‘*The Blue Danube*’ waltzed John around the ballroom. This follows straightforwardly: the causer must be construed as more strongly agentive and in control than the causee, a reading not easily available if the causer is inanimate.

Against alternative analyses: Biggs (2019), although recognizing that English causatives of motion verbs are transitives, argues against the causee receiving a patient θ -role and proposes a new, ‘in motion’ θ -role assigned by a silent verbal affix. Besides inventing an ad-hoc, cross-linguistically otherwise unattested θ -role, the proposal also does not generalize to Hindi-Urdu, Turkish and Sason Arabic which causativize a broader class of unergatives. For Hindi-Urdu, Ramchand (2008) claims that in intransitive unergatives, the sole argument receives both an agent and a patient θ -role, whereas in the causative, it is realized only as a patient, with the agent θ -role assigned to the causer. This proposal fails to explain why only unaccusatives license secondary predicates and cannot account for the inability of unergatives with path arguments to causativize (3b).

References

- Bhatt, R., & Embick, D. (2017). Causative derivations in Hindi-Urdu. *Indian Linguistics*, 78(1–2), 93–151.
- Biggs, A. (2019). Objects in motion verb phrases. *Glossa: a journal of general linguistics*, 4(1), 98.
- Folli, R., & Harley, H. (2006). On the licensing of causatives of directed motion: Waltzing Matilda all over. *Studia Linguistica*, 60(2), 121–155.
- Holisky, D. A. (1987). The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua*, 71, 102–132.
- Ramchand, G. C. (2008). *Verb meaning and the lexicon*. Cambridge UP.
- Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76(4), 859–890.

(Un)intentionally Misleading

Marco Biasio (University of Modena and Reggio Emilia)

The issue. In addressing a number of grammatically relevant reflexes of the basic ontological distinction between accidental and intentional actions, Goncharov (2020a) proposes to relate the constraints on the licenseability of Polarity Sensitive Items (PSIs) in the complement of subject control constructions headed by desire predicates (e.g., English $\llbracket \text{WANT} \rrbracket$) to the nature of the presuppositional content entertained by the modal operator. This change in presuppositional content, formalized as a mechanism of dynamic revision of the attitude holder's beliefs BEL_S , takes place upon the restrictive relativization of the predicate to the set of doxastic worlds determined in the modal base f_{BEL_S} . In particular, strong Negative Polarity Items (NPIs), such as the English minimizer *a damn thing*, and Positive Polarity Items (PPIs), such as the English indefinite *someone*, can be either licensed or anti-licensed in subject control constructions depending on **i**) the interpretation of an action ϕ as intentional or accidental (including cases of intentionally initiated actions which are partially or completely out of the attitude holder's control; cf. the denotations of the English predicates $\llbracket \text{OFFEND} \rrbracket$ or $\llbracket \text{WIN} \rrbracket$) and **ii**) the presence or absence of antimorphic operators (e.g., an overt negation), which impose by default a stative reading of the c-eventuality (Goncharov 2020a: 788–791). Curiously, although Goncharov's (2020a) technical apparatus is tailored to cope with the presuppositional component of certain (doxastic-driven) bouletic attitudes, it is not clear whether the same conclusions can be reached for other linguistic objects of the same class, such as intention reports (e.g., English $\llbracket \text{INTEND} \rrbracket$), which—at least intuitively—should be primarily sensitive to the intentionality parameter. As a matter of fact, at least under ordinary conditions, the semantic computation of desire predicates and intention reports proceeds on parallel tracks in that they interact differently with a number of idiosyncratic properties (Grano 2017: 590–594). The ultimate deviation between the two attitudes boils down to the fact that only intention reports are subject to the same rationality constraints imposed on belief reports. In other words, for intention reports it must always be the case that the attitude holder believes they can see to it that ϕ , as formalized in (1) below (**H** and **G** are strong Priorean temporal operators for the past and the future, respectively). This explains the distinct contradictoriness of (2a), as opposed to the pragmatic oddity of (2b);

- (1) $\llbracket \text{INTEND} \rrbracket := \mathbf{H}(\phi) \wedge \mathbf{G}(\phi) \models \square_{BEL_S}[(\text{STIT} \rightarrow \phi) \wedge (\neg \text{STIT} \rightarrow \neg \phi)]$
(2a) [#]*John intends to fly to the moon, even though he knows this is impossible.*
(2b) *John wants to fly to the moon, even though he knows this is impossible.*

Quite interestingly, however, other parametric differences displayed by intention reports with respect to desire predicates—such as non gradability, upward monotonicity, and lack of conflicting intentions—can be felicitously challenged if the latter are read off not as predicating the attitude holder's psychological desires but, rather, construed as action-relevant effective preferences, i.e., preferences that directly guide and influence the agent's actions (Condoravdi & Lauer 2016: 25–26). The contextual neutralization of the differences between bouletic attitudes taps into the broader issue of whether there are indeed environments where they display a similar sensibility to the intentional parameter. A case study from Contemporary Russian (henceforth CR) is introduced below in support of this hypothesis.

Case study and proposal. It is common knowledge that contemporary Slavic languages impose on almost every form of the verbal paradigm—independently of their tense, mood, and diathesis—a grammatical (i.e., obligatory) marking for viewpoint aspect, mostly as a binary morphosyntactic opposition between perfective (PF) and imperfective (IPF). In CR, as for all the East Slavic languages, the classic denotation of PF, i.e., a function from predicates of events to predicates of times (Kratzer 1998), is enriched by **i**) a scalar implicature SI generating the inference that the eventuality e of type $\langle v, t \rangle$ which is predicated of has started (Goncharov 2020b: 58–59) and **ii**) a conventional presupposition requiring that e is assigned to a uniquely defined point in time as part of a larger sequential chain of qualitatively different eventualities. (Note that IPF, on the other hand, triggers neither requirement.) With an intensional viewpoint operator PF, the syntactic and semantic scope of the antimorphic operator \neg is narrow (i.e., eventive). All the ingredients of the updated denotation are provided in (3):

- (3) $\llbracket \text{PF} \rrbracket^{c,s,w,t} := \lambda P. \lambda t. \lambda w. \exists e_{\langle v, t \rangle} (\tau(e) \subseteq t) \wedge \text{SI}[t' \subset t : \text{START}(e) \circ t'] \wedge \exists t'' [t'' \subset t \wedge \text{END}(e) \circ t'']$

Inasmuch as the categorial status of state-related attitudinal objects (e.g., participant-internal desires and intentions) within the Aristotelian square of linguistically relevant modal oppositions has been questioned for a long time (Moltmann 2020: 172), an association has often been drawn between bouletic attitudes—modulo the variability different doxastic accessibility relations display with respect to homogeneization—and (strong)

deontic worlds. Not only are (strong) deontic and doxastic modal bases conceivable as idealized ontological sets compatible with intentional actions; bouletic attitudes are also taken to entail causal self-referentiality, for they are satisfied if they stand in an appropriate causal relation with an holder’s intentional attitude (Grano 2017: 597). This means that in every optimal bouletic world in which there is an intention report predicating an eventuality which has started, that eventuality (all things being equal) is supposed to reach its endpoint as well. Conversely, in negative modalized utterances, the eventuality licensed in the complement of either bouletic attitude is generally coded as IPF, for the attitude holder sees to it that they do not engage whatsoever in bringing about *e*. See the contrast between the minimal pair (4a) and (4b), where no SI is generated, nor any presupposition of *e*’s temporal uniqueness is entertained. [Crucially, the same pattern obtains if we replace (*ne*) *xoč-u* with intention reports such as (*ne*) *namer-eva-ju-s* ‘I do not mean to’;]

- (4a) *Ja* *xoč-u* *ět-o* *s-dela-t*. (↗ *Ja namerevajus’ èto sdelat*’.)
 I.NOM want-PRS-IND-1SG this-ACC.SG do-INF.PF
 ‘I want to do this.’
- (4b) *Ja* *ne* *xoč-u* *ět-ogo* *dela-t*. (↗ *Ja ne namerevajus’ ètogo delat*’.)
 I.NOM NEG want-PRS-IND-1SG this-GEN.SG do-INF.IPF
 ‘I don’t want to do this.’

This tendency is not met without exceptions and sporadic counterexamples of PF eventualities in the complement of negated bouletic attitudes have been sometimes mentioned in the literature. What (5) suggests is that the attitude holder chose intentionally not to go to the disco to see to it that φ (i.e., her, presumably unpleasant, encounter with Èdik) could not be accidentally brought about (the example is taken from Gusev 2021: 205);

- (5) *Ona* *ne* *po-š-la* *na* *diskotek-u, potomu čto* *ne* *xote-la*
 She.NOM NEG go-PST-IND-F.SG on disco-ACC.SG because NEG want-PST-IND-F.SG
vstreti-t’-sja *s* *Èdik-om*. (not: *vstreč-a-t’-sja*^{IPF})
 meet-INF.PF.REFL with Èdik-INST
 ‘She did not go to the disco because she did not want to run the risk of bumping into Èdik.’

That the aspectual choice in the complement of a negated bouletic attitude is indeed sensitive to the intentionality parameter seems to be confirmed by the CR translation of those intentionally initiated actions addressed by Goncharov (2020a), such as [[OFFEND]], whose outcome cannot be controlled by the attitude holder. Again, the infinitive form is coded PF instead of the canonical IPF;

- (6) *V otvet-Ø* *gospodin-Ø* *Abé* *zaveri-l* *sosed-ej, čto* *«ne*
 In answer-ACC.SG mister-NOM.SG Abe assure-PST-IND-M.SG neighbor-ACC.PL COMP NEG
namer-eva-l-sja *oskorbi-t’* *ix* *čuvstv-a*. (not: *oskorbl’-a-t*^{IPF})
 intend-PST-IND.M.SG.REFL offend-INF.PF their feeling-ACC.PL
 [NKRJa: Ol’ga Kuznecova. Prem’era Japonii obvinili v vosxvalenii militarizma // Kommersant, 2013.12]
 ‘In response, Mr. Abe assured his neighbors that “he did not mean to hurt their feelings”’

The **first** proposal this paper would like to advance is thus to provide a formal derivation of the type of presuppositional sensitivity complements of bouletic attitudes display with respect to the accidental-intentional distinction. The idea is that intentionality effects are epiphenomic upon the aspectual denotation sketched out in (3) and could accordingly be explained within post-Kratzerian models of event semantics as well. This, however, taps into a second correlated issue, because what counts as intentional vs. accidental may not always be given as part of the lexical presuppositions of a certain predicate. The attitude holder may well decide to revise their belief set as if the action in the complement of a bouletic attitude were to be considered by their interlocutor *as* accidental—even in case of prototypically intentional eventualities. (7) presents the act of looting almost as if it had not been direct responsibility of the armed robber;

- (7) *Ona* *takže* *utveržda-et, čto* *mužčin-a* *ne* *namer-eva-l-sja*
 She.NOM too claim-PRS-IND-3SG COMP man-NOM.SG NEG intend-PST-IND-M.SG.REFL
o-grabi-t’ *supermarket-Ø*. (not: *o-grabl’-a-t*^{IPF})
 loot-INF.PF supermarket-ACC.SG
 [NKRJa: V Ispanii mužčina otkryl strel’bu iz ruž’ja v trgovom centre // RIA Novosti, 2017.01]
 ‘She (a previously mentioned female employee, *Author*) also claims that the man did not mean to loot the supermarket’

To account for these and other similar data, the **second** proposal laid out in this paper is to frame such instances of modalized assertions within a dynamic model for the discourse commitment set of each participant, such as Farkas & Bruce’s (2010) ‘Table,’ where the commitment of the attitude holder’s interlocutor is formalized in terms of a (scalar) Trust operator; this operator, when applied to two agents (AG₁, AG₂), a task (**J**) parasitic on

AG₁'s action-relevant effective preferences, and an abstract temporal relation (τ), returns a revised ontological set of beliefs and goals spatiotemporally located (Castelfranchi & Falcone 2000: 807).

SELECTED REFERENCES

- Castelfranchi C. & Falcone R. (2000). Trust and Control: A Dialectic Link. *Applied Artificial Intelligence: An International Journal* 14(8). 799–823. doi: [10.1080/08839510050127560](https://doi.org/10.1080/08839510050127560). • Condoravdi C. & Lauer S. (2016). Anankastic Conditionals Are Just Conditionals. *Semantics and Pragmatics* 9: 8. 1–69. doi: [10.3765/sp.9.8](https://doi.org/10.3765/sp.9.8). • Farkas D. F. & Bruce K. B. (2010). On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1). 81–118. doi: [10.1093/jos/ffp010](https://doi.org/10.1093/jos/ffp010). • Goncharov J. (2020a). Dynamic Presupposition of ‘want’ and Polarity Sensitivity. Rhyne J. & Lamp K. & Dreier N. & Kwon C. (eds.), *Proceedings of SALT 30*. Ithaca, NY: Cornell University. 779–800. doi: [10.3765/salt.v30i0.4823](https://doi.org/10.3765/salt.v30i0.4823). • Goncharov J. (2020b). Whom to oblige? Marušič L. & Mišmaš P. & Žaucer R. (eds.), *Advances in Formal Slavic Linguistics 2017*. Berlin: Language Science Press. 51–73. doi: [10.5281/zenodo.3764847](https://doi.org/10.5281/zenodo.3764847). • Grano T. (2017). The Logic of Intention Reports. *Journal of Semantics* 34(4). 587–632. doi: [10.1093/jos/ffx010](https://doi.org/10.1093/jos/ffx010). • Gusev V.Ju. (2021). Pravila zameny SV na NSV pri modal'nyx glagolax s otricanijem. *Russkij jazyk v naučnom osveščeenii* 41(1). 192–225. doi: [10.31912/rjano-2021.1.8](https://doi.org/10.31912/rjano-2021.1.8). • Kratzer A. (1998). More Structural Analogies between Pronouns and Tenses. Strolovitch D. & Lawson A. (eds.), *Proceedings of SALT 8*. Ithaca: Cornell University. 92–110. doi: [10.3765/salt.v8i0.2808](https://doi.org/10.3765/salt.v8i0.2808). • Moltmann F. (2020). Truthmaker Semantics for Natural Language: Attitude Verbs, Modals and Intensional Transitive Verbs. *Theoretical Linguistics* 46(3-4). 159–200. doi: [10.1515/tl-2020-0010](https://doi.org/10.1515/tl-2020-0010).

Epistemic Resultant Luck and the Appropriate Target of Epistemic Blame

Jordan Myers

University of Houston, MA Student

A burgeoning literature surrounding epistemic blame and its applications in interpersonal epistemic life has arisen at the intersection of epistemology, ethics, and moral responsibility. Four separable questions about epistemic blame can be independently examined: (1) Is there a distinctly epistemic kind of blame? I am convinced there is. When we blame others, there is often a distinctively epistemic *dimension* of our blame. (2) What is the nature of epistemic blame? When we epistemically blame others, what is it that we *do*? There are many competing accounts of epistemic blame. I will speak about “blame” and “epistemic blame” in general; I aim for my arguments to be evaluated independently of what account one is drawn to. (3) What conditions are required for epistemic blameworthiness? This question again mirrors the literature in the moral domain, which has obsessed over the appropriate conditions to be blameworthy. While there are many defensible positions on conditions for blameworthiness, I am concerned with question (4): What is the appropriate target of epistemic blame? By “target of blame,” I mean what we find someone blameworthy *for*; that to which we point and say, “*I blame you for Φ -ing.*” Unlike the variety of responses to the first three questions, the appropriate target of epistemic blame has been monolithically assumed to be *beliefs*. I disagree. I will argue that the appropriate target of epistemic blame must be shifted from beliefs to epistemic character.

It seems deeply inappropriate to blame people for good or bad resultant luck—i.e., luck in “how things turn out.” The essential nature of resultant luck is counterfactual. An outcome was “subject to resultant luck” if:

- (i) there are two people, A and A*,
- (ii) A and A* decide to perform the same action Φ in the same circumstances c ,
- (iii) A and A* do Φ in c ,
- (iv) A’s Φ -ing results in state S and A*’s Φ -ing results in state S*,
- (v) and the fact that A’s decision to Φ created S as opposed to S* and A*’s decision to Φ created S* as opposed to S was the result of factor(s) f that were outside A or A*’s control at the time when deciding to Φ ,

then the states of affairs S and S* created by A and A* are subject to resultant luck. A* need only be a hypothetical agent to claim that the outcome(s) of A’s decision to Φ in c was subject to resultant luck. Any outcome that would have turned out differently due to factor(s) outside one’s control at the time is subject to resultant luck.

The “rationalist” reply to this concern states the appropriate target of blame cannot be subject to resultant luck. The rationalist solution adheres to the control principle, which many find intuitively rational, both initially and after critical reflection. The control principle states that “we are morally assessable only to the extent that what we are assessed for depends on factors under our control.” It seems inescapably inappropriate to blame someone for a hand tremor, a sneeze, or a mental illness which they clearly cannot control. Secondly, the rationalist response differentiates hypological judgements of

blameworthiness with deontic or axiological judgements about whether an act is right or wrong or whether the state created by an action is good or bad, respectively.

The rationalist concludes that it is unfair and inappropriate to blame someone for their resultant luck. From this solution, I extract a principle about the appropriate target of epistemic blame: *That which is subject to epistemic resultant luck (ERL) is an inappropriate target of epistemic blame.*

Epistemic resultant luck mirrors the counterfactual characterization of moral resultant luck: luck in how things turn out, where how things turn out is determined by factors outside the agent's control. Beliefs are subject to ERL. Random factors outside one's control—themselves determined by a myriad of random causes—produce beliefs which are subject to ERL. If any of the relevant causal factors had been different when forming a belief, that belief would be different as well. Because one's beliefs are affected by factors which are outside one's control, targeting epistemic blame at those beliefs is inappropriate.

The *appropriate* target of epistemic blame is one's epistemic character. Epistemic character is roughly the analog of moral character. Moral character refers to the presence or absence of various virtues or vices. Epistemic character is analogous; we can speak about people having poor epistemic characters in the same way they may be said to have poor moral characters. Dogmatism, wishful thinking, hasty reasoning, and certain kinds of biased cognition have been referenced as prime examples of poor epistemic character traits. Boulton adds that being "*intellectually irresponsible, or intellectually vicious, or reckless, or just plain 'stupid'*" are epistemic vices.

Targeting blame at one's epistemic character adheres with our extracted principle; character is not subject to ERL because it is determined *before* one knows how things will turn out—i.e., what beliefs one happens to arrive at. A biased, careless, and lazy thinker is blameworthy for those traits regardless of her good or bad ERL—if she happens to arrive at a true or false belief. Adopting this view entails the striking conclusion that one can be epistemically blameworthy for settling on a true belief if it was done in an epistemically vicious manner.

Resultant luck crucially entails a lack of control over the factors which determine how things turn out. Targeting blame at epistemic character resolves this concern. Our characters are more within the immediate sphere of our control than are the random circumstantial factors which lead us to believe one proposition over another.

When people believe badly, our intuitions about their blameworthiness for their *beliefs* can be better explained by their exuding a poor epistemic character. Believing poorly is only blameworthy when one demonstrates a classic epistemic vice. Instead of targeting blame at beliefs, we should praise or blame others for their virtuous or vicious epistemic characters. We must understand more than simply what people believe to appropriately blame them; we must know who they are; we must know their epistemic character.

Works Cited

- Boult, C. (2020). There is a distinctively epistemic kind of blame. *Philosophy and Phenomenological Research* 103 (3):518-534.
- Boult, C. (2021a). Epistemic blame. *Philosophy Compass* 16 (8):e12762.
- Boult, C. (2021b). The Significance of Epistemic Blame. *Erkenntnis*:1-22.
- Brown, J. (2020a). Epistemically blameworthy belief. *Philosophical Studies* 177 (12):3595-3614.
- Brown, J. (2020b). What is Epistemic Blame? *Noûs* 54 (2):389-407.
- Justin Coates, D. & A. Tognazzini, Neal (eds.) (2013). *Blame: Its Nature and Norms*. New York: Oxford University Press.
- Homiak, M., "Moral Character", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2019/entries/moral-character/>>.
- McHugh, C. (2012). Epistemic deontology and voluntariness. *Erkenntnis*, 77, 65–94.
- Nagel, T. (1979). "Moral Luck." In Nagel, T. *Mortal Questions*. Cambridge University Press.
- Nelkin, Dana K., "Moral Luck", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2021/entries/moral-luck/>>.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Sher, G. (2006). *In praise of blame*. OUP: Oxford.
- Tognazzini, N. and Coates D. J., "Blame", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2021/entries/blame/>>.
- Turri, J., Alfano, M., and Greco, J., "Virtue Epistemology", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/>>.
- Williams, B. (1981). Moral Luck, In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139165860. (pp. 20-39)
- Wolf, S. (1980). Asymmetrical freedom. *Journal of Philosophy* 77 (March):151-66.
- Wolf, S. (1987). Sanity and the Metaphysics of Responsibility. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press. pp. 46-62.
- Wolf, S. (2001). The Moral of Moral Luck. *Philosophic Exchange* 31 (1).
- Wolf, S. (2011). Blame, Italian style. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon* (pp. 332–347). Oxford: OUP.
- Zimmerman, M. J. (1987). Luck and moral responsibility. *Ethics* 97 (2):374-386.
- Zimmerman, M. J. (2002). Taking luck seriously. *Journal of Philosophy* 99 (11):553-576.

Differential offensiveness of ethnophaulisms in L1 vs L2/L3: Non-trivial effects of language choice

Michał B. Paradowski & Marta Gawinkowska
Institute of Applied Linguistics, University of Warsaw
(Institut für Angewandte Linguistik, Universität Warschau)

Recent research (Costa *et al.* 2014; Geipel, Hadjichristidis & Surian 2015, 2016; Cipolletti, McFarlane & Weissglass 2016; Corey *et al.* 2017; Hayakawa *et al.* 2017; Čavar & Tytus 2018; Brouwer 2019; Karataş 2019; Dylman & Champoux-Larsson 2019; Driver 2020) has shown that the same dilemma may elicit different moral judgements depending on the language in which it has been described. For instance, reading scenarios in which noble intentions lead to bad outcomes—or good outcomes ensue despite dubious motives—in a foreign language (L₂) lead participants to place greater weight on outcomes in making moral judgments (Geipel, Hadjichristidis & Surian 2016). This clashes with notion that the effort of using a L₂ cues our cognitive system to prepare for strenuous activity and thus a more deliberate mode of thinking (careful reflection makes people think more about the underlying intentions), and instead suggests that a L₂ reduces the relative weight placed on intentions versus outcomes by either muting emotional responses (triggered by intentions) or depleting cognitive resources. The explanation usually invoked in the above scenarios is the different purported emotionality of content presented in one's first vs second languages.

Using a covert 2×2×2 experiment where 61 L₁ Polish – L₂ English bilinguals – final year MA students majoring in either Applied Linguistics (with English as their first foreign language) or English – were asked to translate (L₁↔L₂) a passage peppered with swearwords, we show that the picture is much more complex (Gawinkowska, Paradowski & Bilewicz, 2013). While the results ostensibly corroborate the so-called 'foreign language effect', with a significant interaction between the source and target words and the direction of translation, it was only observed in the case of *ethnophaulisms*, that is expletives directed at social (out)groups (significant interaction between the source and target words, direction of translation, and type of words, $F(1,59)= 59, p<.01; \eta_p^2=.16$), but not *generic swearwords*. This indicates that the key factor modulating response strength is not so much the different emotional power associated with the respective languages, but *social and cultural norms*.

In a follow-up study, we extend the investigation of the effect of language choice on acceptability judgments of social norm violations by looking at whether a difference will be observed in ratings given in multilingual speakers' L₃ vs L₂. Expectedly, the acceptance rates of scenarios presented in the two languages did not differ much. However, regression analyses indicated different predictors of evaluations: in the L₂, the age factor was significant, with older participants more severe in their judgments; the severity of judgments passed in the L₃ in turn depended on the respondent's gender, with male participants evaluating the scenarios as less severe.

Long cultural learning and socialisation make expressions in L₁ highly prone to normative influences, whereas using a second/third language exempts the speaker from these (whether our own or socially imposed) norms and limitations. It transpires that switching to a foreign language during decision-making may not only reduce emotionally-driven responses and political correctness biases, but also promote candid deliberative processes (e.g. rational cost-benefit considerations).

Reference:

Gawinkowska, M., Paradowski, M.B. & Bilewicz, M. (2013). Second language as an exemptor from sociocultural norms. Emotion-Related Language Choice revisited. *PLoS ONE*, 8(12): e81225. DOI: [10.1371/journal.pone.0081225](https://doi.org/10.1371/journal.pone.0081225)

Sociolinguistic Variation, Speaker Agency, and Local Semantic Conventions

Noah Betz-Richman

Columbia University (PhD Student)

There are many ways to say something. Speakers in a large language population exhibit widespread variation across all domains of language analysis, including phonology, morphology, syntax, and lexical selection. The following pairs are truth-conditionally equivalent, but use different variants (Nowak 2022):

- (1) a. Adonis saw himself in the mirror.
b. Adonis seen hisself in the mirror.
- (2) a. Nobody has heard anything about any festschrift.
b. Ain't nobody heard nothing about no festschrift.

Variants are often associated with particular social groups and identities; this phenomenon is the focus of *variationist sociolinguistics*. For example, the morphosyntactic variants in (1b) and (2b) are associated with working-class speakers and colloquial contexts, as compared to standard forms in (1a) and (2a). Recently, sociolinguistic variation has received greater attention in formal semantics and philosophy of language (see e.g. Nowak 2022, Keiser 2022, Burnett 2019, McCready 2018, Asher & McCready 2014). Theorists aim to explain how classes of variants express social meaning and reflect social identities. This project has both theoretical implications in semantics and sociopolitical importance concerning the relation between language and social identity.

The role of speaker agency in sociolinguistic variation is contested. In this paper, I argue that recent approaches to the semantics of variation overemphasize the role of agency, and conflate instances of literal and sincere conformity to local semantic conventions with a range of pragmatic phenomena. I show that these phenomena are theoretically distinct, and motivate a theory of the semantics of variation in terms of fine-grained, local semantic conventions (cf. Armstrong 2016).

Following Eckert (2012), scholars recognize three 'waves' of theorizing about sociolinguistic variation. These waves reflect different degrees of emphasis on speaker agency in the selection of variants. First-wave approaches minimize the role of agency, and view variation as an unreflective marker of the speaker's social identity; speakers only exercise agency when self-monitoring to suppress variation from standard forms in formal contexts. Second-wave approaches recognize the selection of variants as expressions of the speaker's agency: speakers use vernacular forms to express pride and reflect their membership in local communities. However, these approaches view membership in a social group as a prior, independent, and non-linguistic phenomenon. Third-wave approaches emphasize speaker agency most heavily. Rather than using variants to reflect pre-existing membership in a social category, third-wave approaches understand speakers as actively constructing a social identity in discourse via their selection of variants.

Recent theories of the semantics of variation follow third-wave sociolinguistic approaches in explaining the function of variation in terms of the speaker's ability to construct a nuanced social identity in discourse. Some theorists provide game-theoretic formal models according to which the significance of a variant is its effect on the state of the game. These games model a speaker's efforts to be perceived as possessing various positive attributes such as friendliness and competence (Burnett 2019); or to situate oneself and one's interlocutor into hierarchies of social status (Asher & McCready 2014, McCready 2018). Other theories analyze the significance of variation in terms of illocution: different variants create fine-grained distinctions

between speech-act-types (Nowak 2022). On both sorts of theories, every use of a variant is a strategic, rational move in discourse reflecting the speaker's exercise of linguistic agency.

An empirical motivation for third-wave inspired theories concerns the diversity of functions of variation: while in many cases speakers appear to reflect their own social identities by using variants, speakers also use variants to mock, pejorate, or distance themselves other linguistic sub-communities, to index positive attributes of a distinct sub-community, or to create further divisions within an existing social group (Eckert 2012). The idea is that this diversity of function cannot be explained in terms of pre-existing associations between variants and social groups, and must instead be explained on a case-by-case basis in terms of how the speaker intends to situate herself in the social landscape on a particular occasion of use.

I argue that in their attempts to explain this diversity of function under a single theoretical umbrella, third-wave theories of the semantics of variation conflate theoretically distinct phenomena. In particular, there are important differences between cases in which a speaker sincerely intends to conform to the local conventions of a linguistic sub-community, and cases in which the speaker makes manifest her intention to flout these local conventions. The former cases involve literal usage of a variant, relative to local conventions; the latter cases can be analyzed as a class of Gricean Manner implicatures called ventriloquisms (Nunberg 2018).

In central cases of sociolinguistic variation, speakers reflect their membership in a linguistic sub-community by literally and sincerely conforming to the operative local conventions of that sub-community. However, speakers can also achieve a diverse range of pragmatic effects by manifestly flouting these local conventions. The resulting fine-grained conventionalist theory of sociolinguistic variation better explains the range of empirical phenomena described above.

My theory places weaker emphasis on speaker agency than rival approaches in semantics. Not every use of a variant is an exercise of agency. As conventionalist theories of language emphasize, it is not context-specific speaker intentions that imbue our words with meaning, but prior semantic conventions. I argue that the same is true of sociolinguistic variation. However, semantic conventions must be finely individuated to account for widespread variation between sub-communities. Sociolinguistic variation enables speakers to build fine-grained linguistic sub-communities and to affiliate themselves with these communities.

References:

- Armstrong, Josh. 2016. "The Problem of Lexical Innovation." *Linguistics and Philosophy* 39: 87-118.
- Burnett, Heather. 2019. "Signaling games, sociolinguistic variation, and the construction of style." *Linguistics and Philosophy* 42: 419-450.
- Eckert, Penelope. 2012. "Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation." *Annual Review of Anthropology* 41: 87-100.
- Keiser, Jessica. 2022. *Non-Ideal Foundations of Language*. Oxford: Routledge.
- McCready, Elin. 2018. *The Semantics and Pragmatics of Honorification*. Oxford: Oxford University Press.
- McCready, E. and Nicholas Asher. 2014. "Discourse-Level Politeness and Implicature." In *New Frontiers of Artificial Intelligence*, edited by Y. Nakano, K. Satoh, and D. Bekki.
- Nowak, Ethan. 2022. "Sociolinguistic Variation, Speech Acts, and Discursive Injustice." *The Philosophical Quarterly*.

Nunberg, Geoffrey. 2018. "The Social Lives of Slurs." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss. Oxford: Oxford University Press.

Giles Howdle

University of Edinburgh

g.h.howdle@gmail.com

Abstract: The Linguistic Requirement on Intentional Agency

This paper argues that full-blooded rational agency requires language-possession. But more than this, the possession of language *explains* how rational thought, deliberation, and intentional action are possible. Hence, being a language-user partly *constitutes* being a rational agent. In this paper, I provide an *a priori* argument for this conclusion drawing on Robert Brandom's inferentialism. I also provide an empirical argument drawing on work in the philosophy of cognitive science.

In the first section of the paper, I distinguish the sense of rational agency in which I am interested. While there are a number of interesting lesser senses of agency, by 'full-blooded rational agency', I mean the distinctive form of agency that adult humans typically possess and babies, cats, and caterpillars lack.¹ To have normative practical reasons, for it to be true that you ought to do this, rather than that, requires this form of agency. This is why a normal adult human ought not to torture mice for fun, but the same can't be said for a cat. In short, only full-blooded rational agents are responsive to reasons and subject to rational norms.

Full-blooded agency (and action), as I am calling it, is the kind of agency investigated by Elizabeth Anscombe (2000 [1957]) and Donald Davidson (1980, 1982). Davidson labels it 'intentional agency': the capacity to perform *intentional* actions—actions caused by certain (combinations of) mental states, i.e., intentions. In other words, these mental states *rationalise* action. A key insight of their work is that we can only understand beings as rational agents, and their behaviour as intentional, rational action, if we think they possess 'propositional attitudes', some 'rich pattern of beliefs, desires, and intentions' (Davidson 1982, 318). And we need to understand rational agents and rational action in this way to understand ourselves as e.g., responsible in ways that cats are not. Roughly, the idea is simply that what makes action so different from unintentional behaviour consists in differences in the actors' minds.

In the second section of the paper, I develop an *a priori* argument for the view that language is required for rational agency. My development echoes but differs from Davidson's (1982) famous argument for this conclusion (and I think it avoids some of the problematic features of that argument). I utilise Brandom's inferentialist framework to elucidate the conceptual and mental holism which, I think, explains why intentional agency requires language.

That argument runs roughly as follows. The intentional state of belief, on the inferentialist picture, is understood as a kind of claim one attributes to oneself (or another). It's a conversational commitment on someone's personal deontic scorecard, so to speak. But the belief *that p* can only be understood as a belief *that p* in virtue of its inferential connections with other claims. We wouldn't attribute a belief to someone whose belief lacked these kinds of inferential connections with other commitments. And this kind of grasp on claims' inferential connections seems sufficient to make someone count as a competent language user. Likewise, any intentional mental state such as belief, desire, or intention necessarily requires language-possession for the same reason; propositional attitudes require a degree

¹ The infant may be a 'rational creature' in the sense that it is a human, and it will typically gain full rationality as the kind of being it is. But it does not currently possess full rational agency. See Davidson (1982, 317) for this distinction.

of conceptual mastery which plausibly requires language-possession.² Since propositional attitudes such as belief, desire, and intention are required for rational agency, I conclude that language is required for rational agency.

In the third section of the paper, I provide an empirical argument for my conclusion. I claim full-blooded rational agency requires the capacity for what Andy Clark calls ‘second-order cognitive dynamics’ (Clark 1996, 177), which is roughly thinking about thought. I argue that, in humans at least, our capacity for this kind of agency-enabling second-order cognition is explained by our possession of language.

Language is not simply a means of communication. It is also a transformational mental tool. Like any tool, it allows us to do a lot more than we could otherwise. At least, many philosophers and cognitive scientist have thought so. Daniel Dennett, for instance, claims that becoming initiated into language profoundly reprograms the brain, resulting in entirely new neural machinery and cognitive capacities and phenomena (Dennett 1991 218-219 and 1995, 370-373). It enables a particular form of intentionality, where intentional systems can adopt the ‘intentional stance’ toward themselves. To reflect in the way rational agents can, we need a shift from ‘first-order’ intentional systems to ‘second-order’ intentional systems (Dennett 1995, 121). Animal thoughts may be about things, but their thoughts can’t be about animal thoughts. In other words, reflective thought requires second-order thought.

In other words, second-order thought, thought about thought, requires the object of thought to be another thought. The question is what is necessary for these first-order thoughts and mental states to be receivable by the mind in such that they can become the objects of further thoughts. And the big idea is that the cognitive tool of language is required. By labelling and fixing an idea in a public language, we both ‘offload’ the content of the cognition to the social and inferential realm of language and thereby ‘fix’ its meaning and content. A public language permits us to stabilise cognition into (potentially expressible) linguistically fixed thoughts. And this stabilisation of cognition into thoughts, through the vehicle of public language, is what provides thoughts with the necessary conceptual and inferential stability for them to become the objects of further thought. This linguistic stabilisation of cognition ‘enables us to inspect and criticize our own reasoning in ways that no other representational modality allows’ (Clark 1998, 177). Only through language-possession is this kind of cognition possible for us. Hence, language-possession is partly constitutive of the self-reflective, second-order modes of thought which rational agency requires.

In sum, the paper provides both an a priori and empirical argument for the claim that rational agency requires language. The a priori argument yields the conclusion that this dependency relation is necessary, whereas the empirical argument yields the conclusion that it is merely contingent.

² Non-linguistic beings can be sensibly thought of *as if* they have propositional beliefs and desires, but the sense in which they have beliefs and desires is both lesser and parasitic on the richer propositional, conceptually rich sense of belief and desire that linguistic beings can possess.

References

Anscombe, G.E.M. 2000. *Intention*. Harvard University Press, Cambridge MA.

Clark, A. 1996. Linguistic anchors in the sea of thought? *Pragmatics & Cognition*, 4(1), pp.93-103.
----- 1998. "Magic Words: How Language Augments Human Computation" in Carruthers and J. Boucher (Eds) *Language and Thought: Interdisciplinary Themes* (Cambridge University Press: Cambridge) P.162- 183

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

----- 1982. "Rational Animals", *Dialectica*, 3(4): 317–327.

Dennett, DC., 1991. *Consciousness Explained*. Boston, MA: Little, Brown & Co.

----- 1995. Animal consciousness: What matters and why. *Social Research*, pp.691- 710.

On Moral Responsibility

In his 2020 paper: “From Individual to Collective Responsibility: There and Back Again”, Kirk Ludwig takes up the question of “who is morally responsible for some result ‘*r*’?” in cases where ‘*r*’ is some grievous moral wrong which has been brought about through some form of collective action.¹ Ludwig focuses initially on anthropogenic climate change as his working example of ‘*r*’, since he says that it is a “salient real world example in which a *group* or *collection* of individual agents seems to be, in the first instance, the locus of both causal and moral responsibility” (Ludwig), though he quickly helpfully offers these other examples as well: “stoning, riots, whisper campaigns, racketeering, acid rain, human trafficking, corporate crime and negligence, and the ocean plastic pollution crisis” (Ludwig). Ludwig posits that “there is a strong prima facie case for saying that it is, in the first instance, we together who are *collectively morally responsible* for global warming” (Ludwig), which sets him upon the same path that the view which Russ Shafer-Landau expresses in his 1994 paper: “Vegetarianism, Causation and Ethical Theory”, is on. Shafer-Landau and Ludwig hold that moral inquiry is conducted by asking the following kinds of questions: “who is morally responsible for the cruelty perpetrated on today’s farms” (Shafer-Landau); “who is morally responsible for anthropogenic climate change?” (Sinnott-Armstrong, Ludwig) “who is morally responsible for the theft of the full amount [\$20,000 when 20 people each stole \$1000]?”; and outrightly, “who is morally responsible for the death of Mr. Peabody?”.

Ludwig, along with Shafer-Landau, Sinnott-Armstrong, and anyone else² who asks questions of the above sort, finds themselves embroiled in the difficulties that arise when one tries to assess individual moral responsibility in the face of individual “causal impotence.”³ According to Shafer-Landau, the *Inefficiency Argument*, of which there are two versions, strong and weak, arises as a direct result of this individual causal impotence, and it is these instances of ‘*r*’ that Ludwig et al. are worried about. Shafer-Landau considers the two versions of the *Inefficiency Argument* with respect to causal impotence in light of what kind of change one’s meat purchases can have on the experience of the suffering of NHA on factory farms: “The stronger version says that ordinary meat purchases⁴ make no causal contribution at all to harms suffered by farm animals. The weaker version claims that ordinary purchases have some deleterious impact, but one so small as to be overridden by other factors” (Shafer-Landau).

Ludwig is convinced that we must ask the above kind of questions, but he’s aware that the implications of accepting that an individual is “causally impotent” in the face of some ‘*r*’ which has come about because of the aggregative impact of many individual’s actions, makes it difficult to assess the moral responsibility of the individual members of the group that brought ‘*r*’ about. Since Ludwig is aware that it takes appending the proper amount of moral reproach to an individual’s actions to justify the claim that one ought to avoid performing actions of this kind, he pushes for a filled-out account of individual moral responsibility in the case of even a “massively overdetermined” (Ludwig) grievous ‘*r*’. For, “[i]t is only collectively that we can reverse the processes contributing to it [anthropogenic global warming] or ameliorate the harms it brings. Since we are only collectively and not individually causally responsible—and none of us can do anything about it alone” (Ludwig, Aylsworth). Essentially, Ludwig is looking for a mechanism that can be used to generate the moral force necessary to induce individuals to make relevant changes:

It is not obvious that collective moral responsibility entails anything about individual moral responsibility. If the result is massively overdetermined, why should I be blamed for contributing to a harm or be required to stop? If there is nothing that I can do alone, then why should I have any obligation? Yet if there is no route

¹ In the cases I plan to discuss, it does not matter whether or not this collective action is a cooperative enterprise, or one which actions of a certain type are aggregated and determined to be part of a set of actions performed by others as well.

² Alastair Norcross (2004) and Shelley Kagan (2011) have trafficked in these kinds of questions too.

³ The notion that one’s individual abstinence from or engagement in an activity, has no effect on whether or not a particular result obtains—one is causally impotent with respect to their individual contributions (meat purchases) to the suffering that a pig feels who lived their entire life on a factory farm. My abstaining from buying meat will not have any effect on whether or not that pig or any other pig suffers on a factory farm; the pig will suffer regardless of my individual meat purchases.

⁴ Not the kind of meat purchases that the CEO of McDonald’s would make; but the kind of meat purchases an ordinary citizen would make.

back from collective moral responsibility to individual moral responsibility, collective moral responsibility becomes detached from pressure to alter collective behavior (Ludwig).

While I agree that determining just what it is that has brought about these grievous instances of 'r' is a necessary project, it's not for the sake of determining "who is 'fully morally responsible' for 'r'?" that this project should be undertaken. If we ask questions like: "who is responsible for 'r'?" rather than asking questions like: "why has 'r' come to be, and of those intentional actions which brought about 'r' who performed what?", then we will have in some cases only trivially to moderately interesting answers (e.g., in the case of anthropogenic global warming, we're all (yes all) responsible to some degree for contributing to the problem). If we ask the latter kinds of questions, we will come to possess some more interesting moral information; as answers to these kinds of question will indicate to us which features of the situation that resulted in 'r' coming to be were unintentional,⁵ and therefore not morally relevant, and which features were intentional and therefore, morally relevant. Once we can determine which of the intentional actions that an agent has performed have resulted in some bad 'r', then we can render a judgment about the actual actions performed by the individual; and such judgments will carry with them moral onus enough to justify obligations to abstain from performing actions of those kinds, or to justify assessing certain kinds of reproach or punishment for the performance of actions of that kind. These judgements carry with them sufficient moral weight to render them moral obligation loci, and therefore, action guiding.

I argue that since the nature of moral responsibility is binary (either one is, or one is not morally responsible for performing an act; and that this responsibility is attached only to those acts which are performed intentionally), it is only those acts which we intentionally perform that should be considered as deliberative objects of which we assess their goodness/badness/moral valence. We should therefore be asking questions which are aimed at determining exactly what part an agent played in bringing about the instance of 'r' that we are interested in, rather than questions aimed at determining the degree to which one is morally responsible for it being the case that some 'r' exists. I claim that if we assess one's moral culpability by considering the moral valence of only those acts which one actually intentionally performed, then we will be able to both, avoid dealing with the objection from causal impotence that Ludwig is responding to (as there will be no way in which it could arise), and avoid getting gored on the horns of the false dilemma Ludwig sketches, which looks something like this:

Horn 1: One is *fully* morally responsible for anthropogenic climate change (the suffering of NHA, the death by 1000-cuts of Mr. Peabody), regardless of what actions each member intentionally performed;
or

Horn 2: One is fully morally responsible only for their small contribution (driving a gas guzzler, buying a burger, or inflicting a single paper cut in a 1000-cut-killing activity).

Aylsworth, Tim, and Adam Pham. "Consequentialism, Collective Action, and Causal Impotence." *Ethics, Policy & Environment*, November 25, 2020, 1–14. <https://doi.org/10.1080/21550085.2020.1848191>.

Kagan, Shelley. "Do I Make a Difference?" *Philosophy & Public Affairs* 39, no. 2 (2011): 105–41.

Ludwig, Kirk. "From Individual to Collective Responsibility (PREPRINT)." In *The Routledge Handbook of Collective Responsibility*, edited by Saba Bazargan-Forward and Deborah Tollefsen, 1st ed., 78–93. Routledge, 2020. <https://doi.org/10.4324/9781315107608-8>.

Norcross, Alastair. "Puppies, Pigs, and People: Eating Meat and Marginal Cases." *Philosophical Perspectives* 18, no. 1 (December 2004): 229–45. <https://doi.org/10.1111/j.1520-8583.2004.00027.x>.

Shafer-Landau, Russ. "Vegetarianism, Causation and Ethical Theory." *Public Affairs Quarterly* 8, no. 1 (1994): 85–100.

Sinnott-Armstrong, Walter, and Richard B. Howarth, eds. *Perspectives on Climate Change: Science, Economics, Politics, Ethics*. Vol. 5. Advances in the Economics of Environmental Resources. Emerald Group Publishing Limited, 2005. [https://doi.org/10.1016/S1569-3740\(2005\)5](https://doi.org/10.1016/S1569-3740(2005)5).

⁵ either non-human, or accidentally caused and therefore not morally relevant.