

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

Masaharu Mizumoto\* and Zhong Yuanyan†

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

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

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

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

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

---

\* Japan Advanced Institute of Science and Technology

† Japan Advanced Institute of Science and Technology

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

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

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

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

## References

- Knobe, J. (2016). Experimental philosophy is cognitive science. In Sytsma, J., & Buckwalter, W. (Eds.). (2016): 78-96.
- Mizumoto, M. (2018a). A simple linguistic approach to the Knobe effect, or the Knobe effect without any vignette. *Philosophical Studies*, 1-18.
- Mizumoto, M. (2018b). “Know” and Japanese Counterparts: “Shitte-iru” and “Wakatte-iru”. In M. Mizumoto, S. Stich, and E. McCready (eds.), *Epistemology for the Rest of the World*. Oxford: OUP. 77-122.