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

Masaharu Mizumoto and Zhong Yuanyan
(Japan Advanced Institute of Science and Technology)

Abstract

Mizumoto (2018a) demonstrated that the moral asymmetry of intention attribution in the Knobe effect can be produced without any vignette, merely through people's felicity judgments about sentences with different moral valence. This apparently suggests that the moral asymmetry is linguistically encoded. In this study, following Mizumoto's approach, we investigate the possibility of producing the opposite moral asymmetry, using two Chinese counterparts to "intentionally". We will also conducted the surveys with the standard Chairman case to see the linguistic effect on the Knobe effect.

Introduction

To what extent language plays a role in producing the Knobe effect? There is at least some role for language to play, according to Mizumoto (2018a). There he produced the analogous moral asymmetry of intention attribution without using any vignette.

There 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. See for example, in the case of harming/improving the environment:

<From Mizumoto (2018a)>
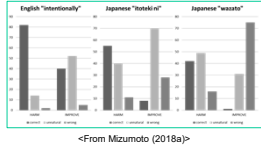
However, as in the figure above, there he also used two Japanese counterparts of "intentionally", which showed the same moral asymmetry, with significantly sharper asymmetries than the one found for English "intentionally".

Given this cross-linguistic diversity, one may 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 a morally bad (intentional) action would be judged unnatural or even grammatically wrong, by the native speakers.

Can there be any such adverb in natural language? In this paper, we looked for the adverb in Chinese language and found one.

Study 1

Methods:

There are several counterparts of English "intentionally" in Chinese.

Among them, the common are the following two:

1. Gu yi de;故意地、
2. You yi de;有意地

1 can be understood as a Chinese counterpart of Japanese "wazato", with the same negative implication.
2 is the standard translation of English "intentionally".

Note that, In **LONGMAN Dictionary of Contemporary English**, "intentional" means "done deliberately and usually intended to cause harm" in English. If this is correct, the negative connotation is already encoded in English "intentionally".

According to widely used dictionaries in China, the standard entries of "intentionally" are 故意地Gu yi de, 有意地You yi de, and 蓄意地 in Chinese. In **Oxford Advanced Learner's English-Chinese Dictionary**, the entries of "intentionally" are again "故意地Gu yi de 、有意地You yi de 、and 存心地". In **THE NEW ENGLISH-CHINESE CHINESE-ENGLISH DICTIONARY**, the entry of "intentionally" is "有意地You yi de ".

In **Modern Chinese Dictionary**, "故意地" is used as an adverb, which means that somebody does something consciously. In the field of law, it is also used as a noun. "故意地" mainly has a negative connotation in Chinese, more commonly used in negative scenarios. Thus, it is used in criminal law in Chinese, to express, for example, "intentional crime" and "Intentionally commits homicide."

When"有意" is used as an adverb, it means deliberately. And when it is used as a verb, it means having a desire to do something. Contrary to "故意地", "有意地" can have both a positive and a negative connotation.

We conducted the analogous surveys in Chinese following Mizumoto (2018)'s method, with these two adverbs.

In particular, participants were presented the Chinese translations of the following sentences:

<Morally neutral>
TUMBLE: X intentionally tumbled.
BREAK: X intentionally broke the vase.
IGNORE: X intentionally ignored Y.
<Morally bad>
HARM: X intentionally harmed the environment.
STEAL: X intentionally stole the purse.
KILL: X intentionally killed Y.
<Morally good>
IMPROVE: X intentionally improved the environment.
SAVE: X intentionally saved the life of Y.

(BREAK and IGNORE are considered morally neutral because they are only morally bad when done intentionally.)

In the actual Chinese sentences used in the surveys, "X" and "Y" were replaced by specific Chinese names, as in Mizumoto (2018a).

Participants were then asked to judge the use of the adverb in these sentences, by choosing among: 1) correct and **natural**, 2) not wrong but **unnatural**, and 3) **wrong**, as in Mizumoto (2018a).

115 participants for 1. Gu yi de (age M = 29.8, 42 males, 73 females), 105 participants for 2. You yi de (age M = 29.6, 53 males, 52 females), were recruited via 51调查网, a Chinese online survey service.

Based on these linguistic considerations, our hypothesis here is that language plays a crucial role in producing the Knobe effect, which is shown by the fact that the moral asymmetry is linguistically encoded in "intentionally" or its counterparts in other languages. Indeed, pace Knobe (2016), the Knobe effect is a linguistic effect of the concept of intentional action captured by a particular language, if the moral asymmetry can be reversed by the linguistic difference.

We investigate this hypothesis by examining whether there is intra-linguistic variance in people's felicity judgments between 1. Gu yi de and 2. You yi de.

In particular, as a Chinese counterpart of Japanese "wazato", with its negative connotation, 1 will show a more acute moral asymmetry than the one found in "intentionally," and having both positive and negative connotations, 2 will show generally high felicity rates, without an explicit moral asymmetry, or even with the opposite moral asymmetry, where sentences about a morally bad action will be judged "unnatural" or "wrong", while sentences about a morally good action judged mostly "natural".

If we find such reversed moral asymmetry in the result of the survey with 2, You yi de, the results of this study are consistent with our hypothesis. The result of the survey with You yi de would even open the possibility that the Knobe effect had actually been largely a linguistic effect, rather than psychological one, contrary to what Knobe (2016) assumed.

Results:

We obtained the results that are consistent with the hypothesis.

<Result of 1. Gu yi de>

<Result of 2. You yi de>
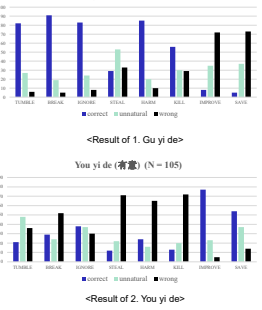
In particular, the result of 1. Gu yi de showed a pattern very similar to the one observed in "wazato" in Mizumoto (2018a) with an exception of STEAL, and the result of 2. You yi de demonstrated the exact opposite of the moral asymmetry of "wazato" and "Gu yi de", where most participants judged the sentences about a morally bad action "wrong", while those about a morally good action are judged "natural".

In particular, see the contrast of HARM and IMPROVE in the results of these two adverbs.
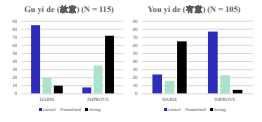
The next question is whether this radical linguistic difference affects the result of the standard Chairman case.

Study 2

Methods:

We conducted a survey asking Chinese participants whether to attribute intention to the agent in the standard Chairman case, using You yi de (有意地).

We used the same online platform (51调查网) as in Study 1, and participants were divided into the HARM group and the IMPROVE group. After eliminating participants who answered both the HARM question and the IMPROVE question, we had 52 participants for HARM (age M = 28.4, 24 males and 28 females), and 76 participants for IMPROVE (age M = 28.9, 33 males and 43 females).

Result:

Against our expectation, the result showed no less acute moral asymmetry as the original Chairman case with "intentionally",
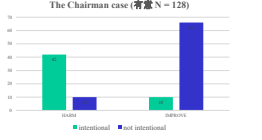
where more than 80% of participants answered that the chairman intentionally harmed the environment, while only 13% answered that the chairman intentionally improved the environment. The effect size (φ) was 0.68, which is huge (where 0.1 is small, 0.3 is medium, and 0.5 is large), and no smaller than that of the original Knobe effect.

General Discussion

Unfortunately, or fortunately for Knobe, what we observed was exactly the same pattern observed in the standard Chairman case for English speakers using "intentionally".

Thus, such a result provides crucial evidence that the role of the linguistic factor in the Knobe effect is limited. Even though the felicity judgments of sentences with You yi de showed the opposite moral asymmetry, that effect was totally 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.

But to what extent this conclusion can be generalized to other (possible and actual) cases of apparent psychological effect?

For each effect found in the fully contextualized use of the relevant terms/phrases, 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 is wholly linguistic. For example, Japanese has two distinct verbs for propositional knowledge, shitte-iru and wakatte-iru, whose uses or judgments about them can radically diverge in some contexts. Thus, the judgments about whether an agent knows some fact or not can differ radically in epistemologically interesting cases (Mizumoto 2018b).

<From Mizumoto (2018b)>
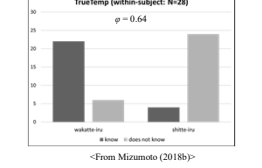
But if so, since they are judgments by the same people about the same situation, such an intra-linguistic diversity is a linguistic effect.

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.

Conclusion

Our studies here have shown a limit of the role of language in producing the Knobe effect, or the moral asymmetry of intention attribution.

However, since we also have clear cases in which a large effect is wholly linguistic, we cannot generalize the present results to other cases without looking at the details, preferably with a further empirical study.

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, sometimes using the approach of Mizumoto (2018a).

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.