

Special Session II: Presentation 2

**Language of Emotions for Simulating Moral Imagination**

Rafal Rzepka<sup>1</sup> and Kenji Araki<sup>2</sup>

Hokkaido University, JAPAN

[kabura@media.eng.hokudai.ac.jp](mailto:kabura@media.eng.hokudai.ac.jp)<sup>1</sup>; [araki@media.eng.hokudai.ac.jp](mailto:araki@media.eng.hokudai.ac.jp)<sup>2</sup>

Cognition processes of a moral judgment have been puzzling researchers for decades and there are different interpretations of our ethical behaviors. There are two influential schools of how our morality works - one represented by Plato and Kant who taught that ethical judgment is a rational enterprise which appreciates abstract reasons for providing direction and motivation, and second one represented by "sentimentalist" philosophers such as Hume and Smith, who were convinced that emotions are the primary basis for moral decisions. Modern scientists, like Joshua D. Greene from Harvard University, believe that emotion and reason both play critical roles and are trying to measure how these roles interact. However, tests on human subjects (especially observations that require technologies like fMRI) are costly and time-consuming. To address this problem, we introduce our idea of using emotive expressions to discover moral consequences of a given action. In our approach we rely on a hypothesis saying that majority of people are sane and make ethically correct decisions, however anyone can use our method for examining his or her own assumptions about human behavior. Automatizing the retrieval of emotional reactions can support traditional survey-based approaches and develop a novel method for investigating big numbers of anonymous and unbiased subjects without actually contacting them, which can become a new opportunity for cognitive scientists and even philosophers willing to confirm their theories. In order to confirm applicability of our idea we have experimented with different kinds of emotionally loaded words and expressions in Japanese language and confirmed their validity for the moral consequences retrieval task by applying them to an algorithm for searching Yahoo Japan service. To measure the usefulness of emotion categories borrowed from Nakamura's "Emotional Expressions Dictionary", which is currently the core of our emotions recognition tool, we have performed two kinds of experiments. First, we created a consequence data set with expressions stating directly what can happen after a given action (praise/punishment, social approbation/lack of social approbation, award/imprisonment, etc.) and implemented it into a program that recognizes "IF-THEN sentences" and counts how many times someone described a consequence of an input action with a non-emotional consequence expression. Then we used the emotional consequence expressions for the same set of input, which is a set of one hundred actions that are ethically questionable or of no remarkable moral consequence, for example "killing a man" vs. "killing a virus". Experimental results show that consequences expressed by emotional categories were matched more often (86% of cases) than non-emotional direct consequences expressions (68% of cases) proving that further development of textual emotion recognition can lead to interesting and more trustful usages of our approach. It is worth notifying that by creating a categorized emotional expression dataset in any language, one can implement "IF-THEN sentences" extracting algorithm and perform various experiments on how different cultures react to a given set of ethically marked or neutral actions. Also different types of experiments (e.g. emotional consequences of emotional actions) become possible.