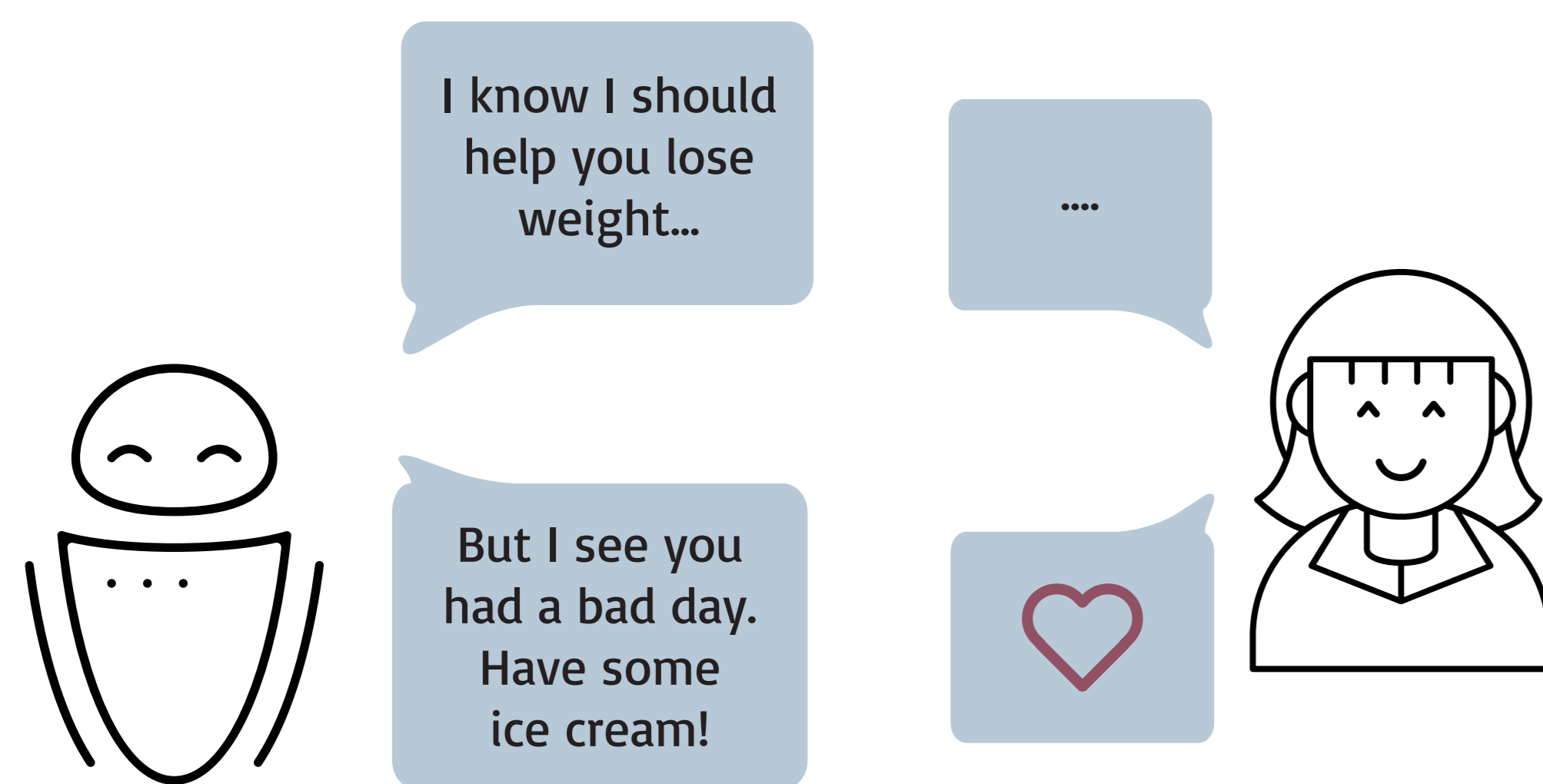


Human values motivate our opinions and actions

Engineering value-sensitive agents that align their actions with human values is essential for robust and **beneficial AI**.



General Values

Broad and abstract

Applicable across contexts

Suitable for societal questions

Context-Specific Values

Applicable to a context

Defined within a context

Suitable for concrete usage

Contributions

Axies: a **methodology** for identifying context-specific values

A collaborative **web platform** to support Axies

An **evaluation** of Axies via a user study involving 80 subjects

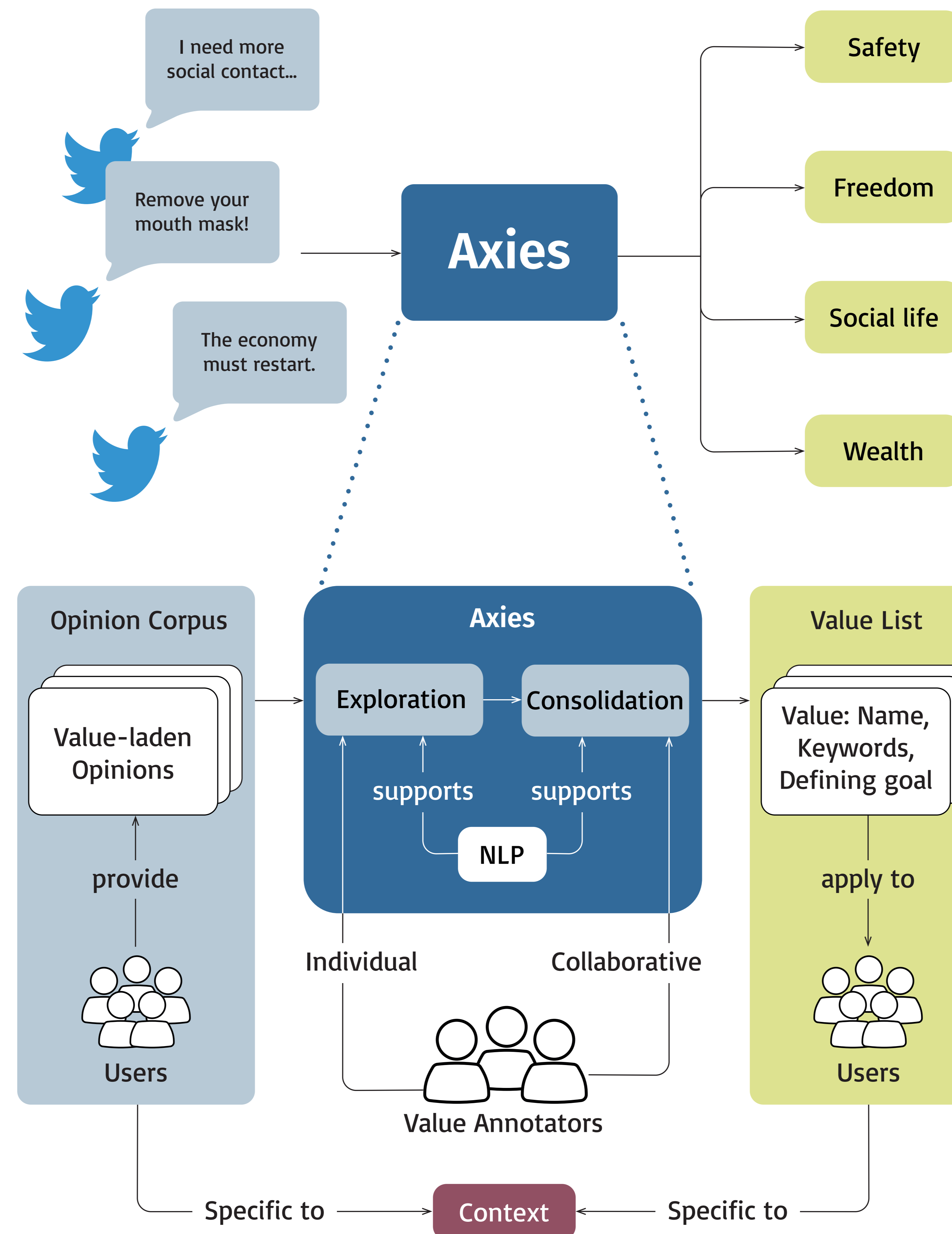
A **comparison** of general and context-specific values

Enrico Liscio, Michiel van der Meer, Luciano C. Siebert,
Catholijn M. Jonker, Pradeep K. Murukannaiah



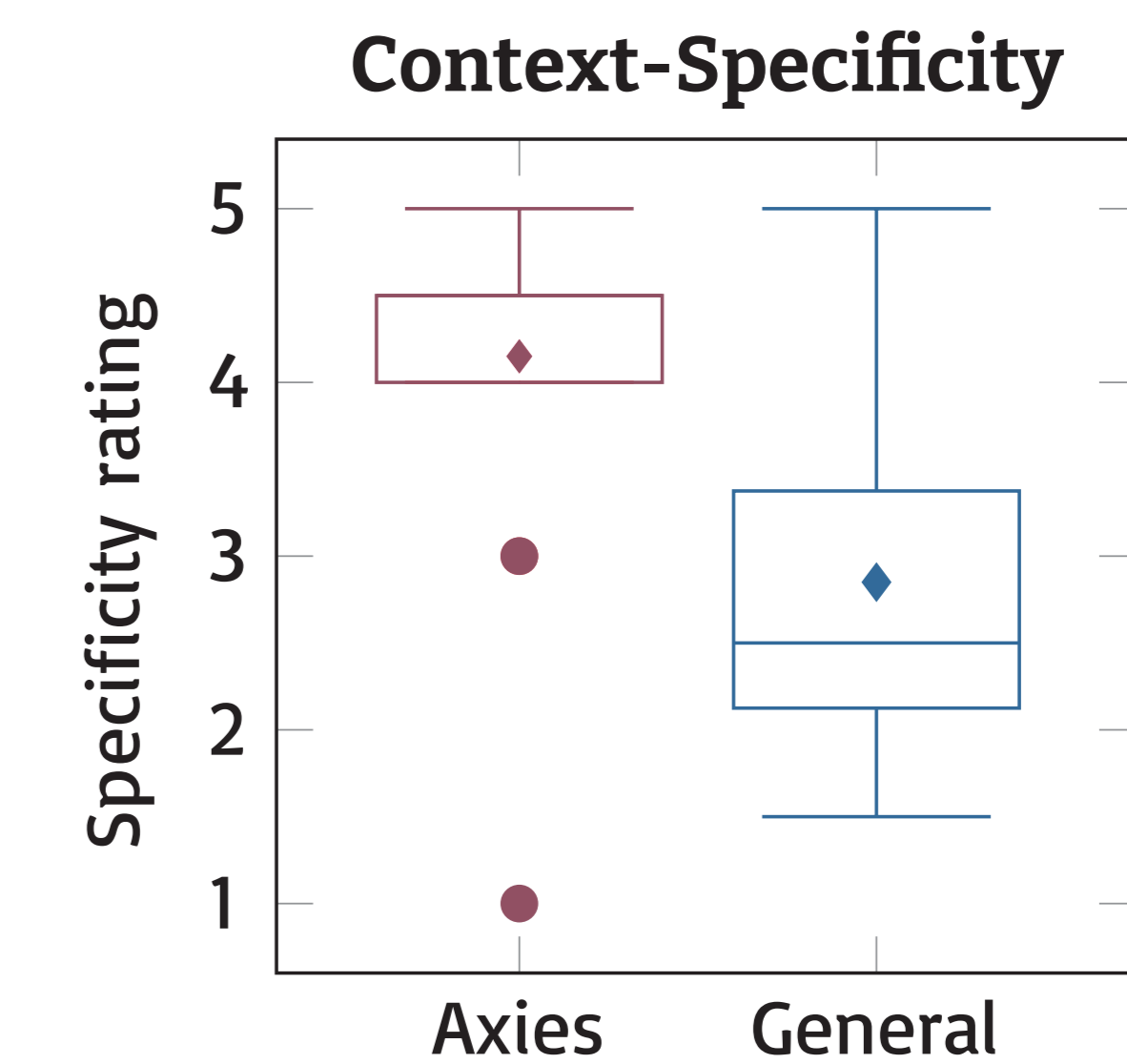
What Values should an Artificial Agent Align with?

A Hybrid Methodology for Identifying Context-Specific Values



Context-specific values are better for concrete tasks

Axies values are **more context-specific** than general values.



Axies values are **annotated more often** and with **higher annotator agreement** than general values.

