

Supplementary Information

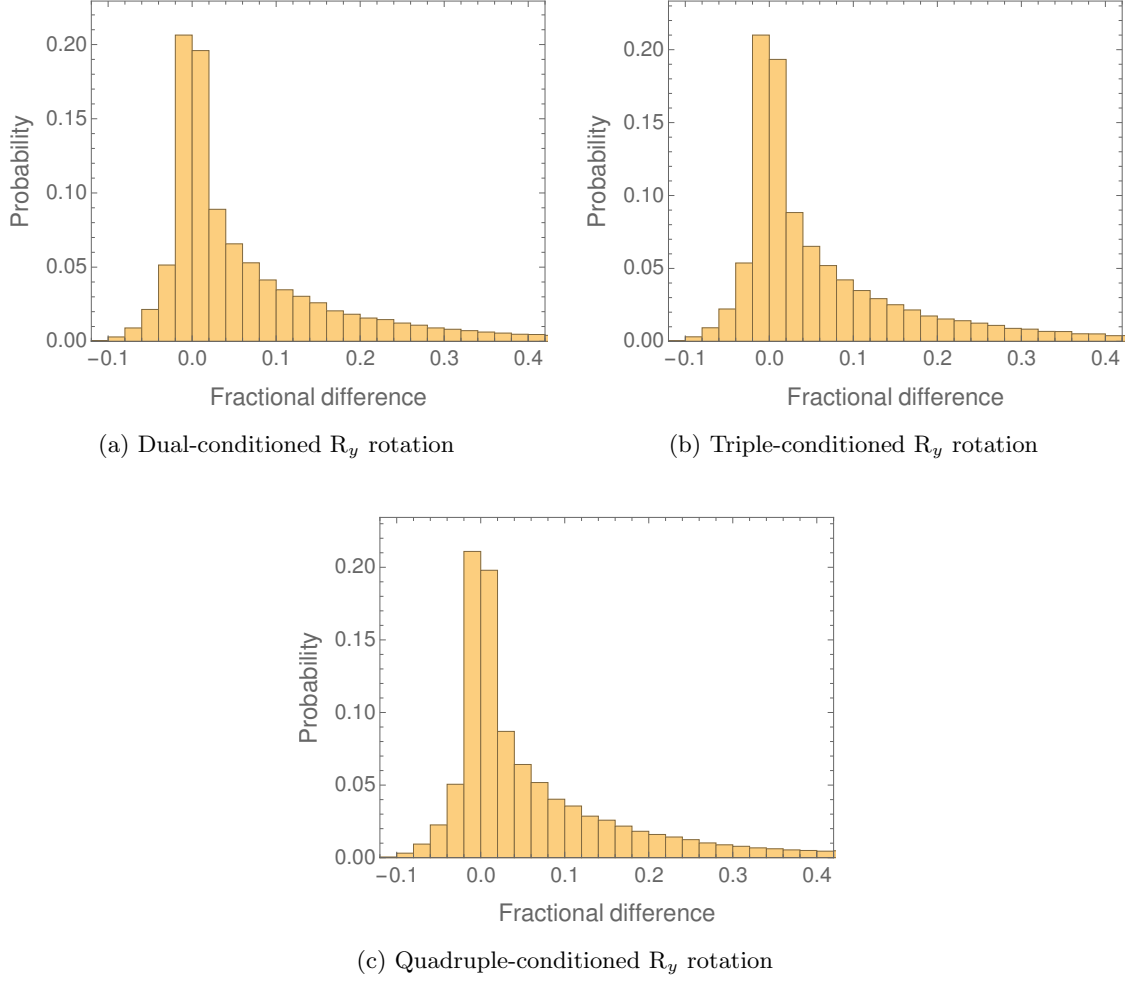
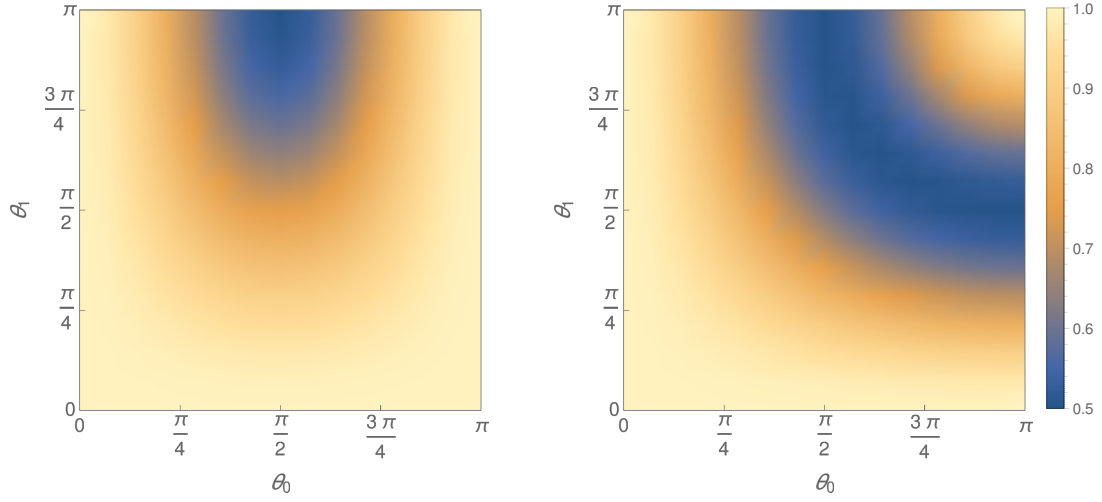


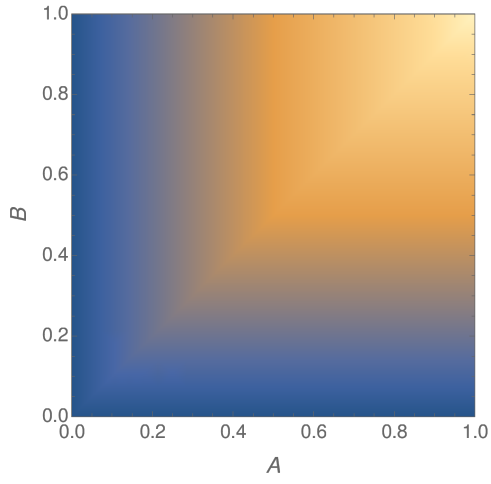
Figure S1: Simulated fractional (relative) difference between the weighed mean (or inner product) determined by the classical fuzzy-based Silicon Coppélia system and quantum logics. The scenario in Eq. (S11) is taken in this simulation. The quantum circuit consists of successive single-control C- R_y rotations, as depicted in Eq. (3) and Fig. 3h. Each condition is made up of one distinct weight factor $|\phi_i\rangle = \cos \frac{\phi_i}{2} |0\rangle + \sin \frac{\phi_i}{2} |1\rangle$, where $\phi_i \in [0, \pi]$ and $\sum_i \phi_i = \pi$. The angle of rotation $\alpha_i \pi$ for each condition is arbitrary with $\alpha_i \in [0, \pi]$ and $\sum_i \alpha_i = 1$. 50,000 combinations of ϕ_i and α_i were randomly generated to evaluate the resultant target state $|\psi\rangle$. The probability of observing state $|1\rangle$ is compared to its classical counterpart, which is given by $\sum_i \sin^2 \frac{\alpha_i \pi}{2} \sin^2 \frac{\phi_i}{2}$. A positive relative difference implies a greater value obtained by quantum logics. A single-conditioned C- R_y rotation gives exact results. The distributions for C- R_y rotations under 2–4 control qubits are similar, with a high proportion showing little deviation.



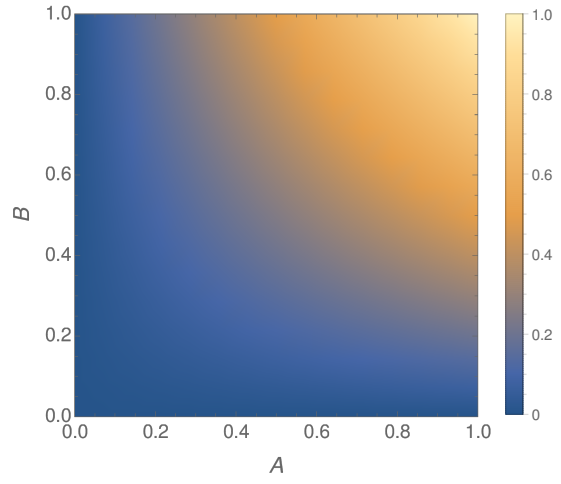
(a) Rotation formalism (Fig. 3f) with $\theta_1 = \alpha\pi$

(b) Entanglement formalism (Fig. 3g) with $|\alpha\rangle = \cos \frac{\theta_0}{2} |0\rangle + \sin \frac{\theta_0}{2} |1\rangle$

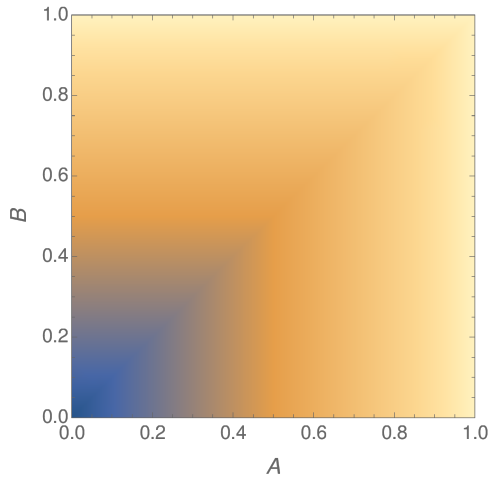
Figure S2: The purity of the target state $\gamma = \text{Tr}(\hat{\rho}^2)$ under the two-state generation formalisms with an additional control qubit $\cos \frac{\theta_0}{2} |0\rangle + \sin \frac{\theta_0}{2} |1\rangle$. $\gamma = 0.5$ represents a fully mixed state.



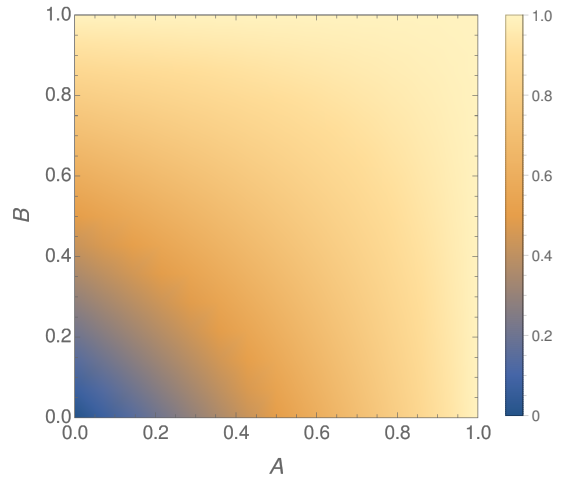
(a) Zadeh (fuzzy) AND



(b) Quantum AND



(c) Zadeh (fuzzy) OR



(d) Quantum OR

Figure S3: The output of the AND (a,b) and OR (c,d) gates using the conventional Zadeh operators for fuzzy logics (a,c) or the quantum logic operators (b,d). For the quantum operators, the values represent the probability of $|1\rangle$.

S1 Silicon Coppélia and Terminology

S1.1 Encoding

During *encoding* the features of a human or virtual agency, each feature is numbered k in a set $\mathcal{F}$, for their level of *ethics* (good $\mathcal{E}^+$ or bad $\mathcal{E}^-$), *aesthetics* (beautiful $\mathcal{S}^+$ or ugly $\mathcal{S}^-$), and *epistemics* (realistic $\mathcal{P}^+$ or unrealistic $\mathcal{P}^-$). During the encoding, moreover, the robot evaluates in how far the user has certain *affordances* (aids $\mathcal{F}^+$ or obstacles $\mathcal{F}^-$), action possibilities that make the other agency functional as a tool to achieve robot goals or not (e.g., to recharge its battery, to keep it from falling, to serve its user). These dimensions constitute a set of appraisal variables $\mathcal{D} = \{\mu = \mathcal{E}, \mathcal{F}, \mathcal{S}, \mathcal{P}\} \times \{\nu = +, -\} = \{\mu^\nu\}$. Features receive different feature weights $w^{(k)}$ to represent subjective judgments of any kind (weights are unconstrained). Each appraisal variable also has its appraisal weight for the feature $d_l^{(k)}$ ($l = \mu^\nu \in \mathcal{D}$) as an indicator of its dominance in the appraisal process. Empirical research found that features pertaining to ethics and affordances are better predictors of use intentions and engagement than epistemic aspects (e.g., the agency is real or virtual) or aesthetics, i.e., $d_{\mathcal{E}^\pm}^{(k)}, d_{\mathcal{F}^\pm}^{(k)} > d_{\mathcal{P}^\pm}^{(k)}, d_{\mathcal{S}^\pm}^{(k)}$.^[1] The encoding process generates the perceived weight of the appraisal variables

$$p_l^{(k)} = w^{(k)} d_l^{(k)}, \quad l = \mu^\nu \in \mathcal{D}. \quad (\text{S1})$$

Table S1 summarizes the parameters used in the encoding process.

Table S1: Variables involved in feature encoding of the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system.

Name	Si-Coppélia notation	Description	Qubit notation
Set of features	$\mathcal{F}$	$\mathcal{F} = \{k = 1, \dots, N_k\} \subseteq \Omega$	
Set of all possible features	Ω	$\mathcal{F} \subseteq \Omega$	
Feature	k	$k \in \mathcal{F}$	

Table S1: Variables involved in feature encoding of the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system. (Cont'd)

Name	Si-Coppélia notation	Description	Qubit notation
Set of appraisal variable labels	$\mathcal{D}$	$\mathcal{D} = \{\mu = \mathcal{E}(eth), \mathcal{F}(aff), \mathcal{S}(aest), \mathcal{P}(ep)\} \times \{\nu = +, -\}$ $= \{\mu^\nu\}$ $\mathcal{E}^+ = good, \mathcal{E}^- = bad,$ $\mathcal{F}^+ = aid, \mathcal{F}^- = obstacle,$ $\mathcal{S}^+ = beautiful, \mathcal{S}^- = ugly,$ $\mathcal{P}^+ = real, \mathcal{P}^- = unreal$	
Appraisal weight	$\vec{d}$	$\vec{d} = (\vec{d}^{(k)})_{k \in \mathcal{F}}$	$ d_{\mathcal{E}^+}^{(k)}\rangle, d_{\mathcal{E}^-}^{(k)}\rangle,$ $ d_{\mathcal{F}^+}^{(k)}\rangle, d_{\mathcal{F}^-}^{(k)}\rangle,$ $ d_{\mathcal{S}^+}^{(k)}\rangle, d_{\mathcal{S}^-}^{(k)}\rangle,$ $ d_{\mathcal{P}^+}^{(k)}\rangle, d_{\mathcal{P}^-}^{(k)}\rangle$
Appraisal weight for feature k	$\vec{d}^{(k)}$	$\vec{d}^{(k)} = (d_l^{(k)})_{l=\mu^\nu \in \mathcal{D}}$ $d_l^{(k)} \in [0, 1]$	
Feature weight	$\vec{w}$	$\vec{w} = (w^{(k)})_{k \in \mathcal{F}}, w^{(k)} \in [0, 1]$	$ w^{(k)}\rangle$
Perceived weight of appraised variable	$\vec{p}$	$\vec{p} = (\vec{p}^{(k)})_{k \in \mathcal{F}}$ $\vec{p}^{(k)} = w^{(k)} \vec{d}^{(k)}$ $= (p_l^{(k)})_{l \in \mathcal{D}} = (w^{(k)} \vec{d}^{(k)})_{l \in \mathcal{D}}$ $p_l^{(k)} \in [0, 1]$	$ p_{\mathcal{E}^+}^{(k)}\rangle, p_{\mathcal{E}^-}^{(k)}\rangle,$ $ p_{\mathcal{F}^+}^{(k)}\rangle, p_{\mathcal{F}^-}^{(k)}\rangle,$ $ p_{\mathcal{S}^+}^{(k)}\rangle, p_{\mathcal{S}^-}^{(k)}\rangle,$ $ p_{\mathcal{P}^+}^{(k)}\rangle, p_{\mathcal{P}^-}^{(k)}\rangle$

S1.2 Comparison

In the *comparison* phase, the features are judged for *relevance* to robot goals j in a set of goals $\mathcal{G}$ (relevant ($R^{(k)}$) or irrelevant ($R^{\dagger(k)}$)) and *valence* to robot goals (positive or negative outcome expectancies ($V^{(k)}$ or $V^{\dagger(k)}$)) using the membership functions. Relevance determines the intensity of the affective response, whereas valence determines its direction. The variables involved in the comparison phase are summarized in Table S2. An agency's features (e.g., its intentions) encoded as positive (e.g., "good") may afford the facilitation of a desired robot goal, for example, to have maintenance every once in a while. This instigates positive outcome expectancies. When the features are found to be highly relevant to the robot goals (for either positive or negative outcome expectancies), a high level of relevance results.

Table S2: Variables involved in the relevance and valence evaluation as well as their related action tendencies in the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system.

Name	Si-Coppélia notation	Description	Qubit notation
Set of choices of action	$\mathcal{A}^{(k)}$	$\mathcal{A}^{(k)} = \{i^{(k)}\}$	
Choice of action for feature k	$i^{(k)}$	$t(i^{(k)}) \in \mathcal{T}$	
Action tendency	$\mathcal{T}$	$\mathcal{T} = \{positive, negative, change, avoid\}$	
Set of goals	$\mathcal{G}$	$\mathcal{G} = \{j = 1, ..., N_j\}$	
Goal	j	$j \in \mathcal{G}$	
Ambition towards goal	$\vec{a}$	$\vec{a} = (a_j \in [-1, 1])_{j \in \mathcal{G}}$ $a_j > 0 \Rightarrow desired$ $a_j < 0 \Rightarrow undesired$	$ a_j^+\rangle$: <i>desired</i> $ a_j^-\rangle$: <i>undesired</i>
Belief for feature k	$\vec{b}^{(k)}$	$\vec{b}^{(k)} = (\vec{b}_j^{(k)})_{j \in \mathcal{G}}$ $\vec{b}_j^{(k)} = (b_{ij}^{(k)} \in [-1, 1])_{i \in \mathcal{A}^{(k)}}$ $b_{ij}^{(k)} > 0 \Rightarrow facilitate$ $b_{ij}^{(k)} < 0 \Rightarrow inhibit$	$ b_{i+j}^{(k)}\rangle$: <i>facilitate</i> $ b_{i-j}^{(k)}\rangle$: <i>inhibit</i>
Membership function of x belonging to $\mathcal{Y}$	$\mu_{\mathcal{Y}}(x)$	$\mu_{\mathcal{Y}}(x) \in [0, 1]$	
Affect (goal j by action i)	$\mu_A(b_{ij}^{(k)})$	<i>affects</i> = <i>facilitates</i> OR <i>inhibits</i>	$ A_{ij}^{(k)}\rangle$
Important (for goal j)	$\mu_I(a_j)$	<i>important</i> = <i>desired</i> OR <i>undesired</i>	$ I_j\rangle$
Agreement to any action for goal j for feature k (for relevance)	$g_j^{(k)}$	$g_j^{(k)} \in [-1, 1]$ $g_j^{(k)} > 0 \Rightarrow agree$ $g_j^{(k)} < 0 \Rightarrow disagree$	$ g_j^{(k)}\rangle$: <i>agree</i> $ g_j^{\dagger(k)}\rangle$: <i>disagree</i>
Relevance of encoded features	$R_p^{(k)}$	$R_p^{(k)} = \mu_R(\vec{p}^{(k)})$ $R_p^{(k)} \in [0, 1]$	$ R_p^{(k)}\rangle$

Table S2: Variables involved in the relevance and valence evaluation as well as their related action tendencies in the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system. (Cont'd)

Name	Si-Coppélia notation	Description	Qubit notation
Relevance & irrelevance from goal comparison	Relevance $\vec{R}^{(k)} = (R_j^{(k)})$, Irrelevance $\vec{R}^{\dagger(k)} = (R_j^{\dagger(k)})$	$R_j^{(k)} = \mu_R(\vec{b}^{(k)}, a_j, g_j^{(k)})$, $R_j^{\dagger(k)} = \mu_{R^\dagger}(\vec{b}^{(k)}, a_j, g_j^{(k)})$, $R_j^{(k)}, R_j^{\dagger(k)} \in [0, 1]$	$ R_j^{(k)}\rangle$: <i>relevance</i> $ R_j^{\dagger(k)}\rangle$: <i>irrelevance</i>
Relevance & irrelevance	$R^{(k)}$: Relevance $R^{\dagger(k)}$: Irrelevance	$R^{(k)} = \mu_R(\vec{R}^{(k)}, R_p^{(k)})$, $R^{\dagger(k)} = \mu_{R^\dagger}(\vec{R}^{(k)}, R_p^{(k)})$, $R^{(k)}, R^{\dagger(k)} \in [0, 1]$	$ R^{(k)}\rangle$: <i>relevance</i> $ R^{\dagger(k)}\rangle$: <i>irrelevance</i>
Agreement to action i for goal j for feature k (for valence)	$g_{ij}^{\prime(k)}$	$\vec{g}^{(k)} = (g_{ij}^{\prime(k)})_{i \in \mathcal{A}^{(k)}, j \in \mathcal{G}}$ $g_{ij}^{\prime(k)} \in [-1, 1]$ $g_{ij}^{\prime(k)} > 0 \Rightarrow agree$ $g_{ij}^{\prime(k)} < 0 \Rightarrow disagree$	$ g_{ij}^{\prime(k)}\rangle$: <i>agree</i> $ g_{ij}^{\dagger(k)}\rangle$: <i>disagree</i>
Valence	$V^{(k)}$ (+), $V^{\dagger(k)}$ (−)	$V^{(k)} = \mu_V(p_{\mathcal{E}^\pm}^{(k)}, p_{\mathcal{F}^\pm}^{(k)}, \vec{b}^{(k)}, \vec{a}, \vec{g}^{(k)})$, $V^{\dagger(k)} = \mu_{V^\dagger}(p_{\mathcal{E}^\pm}^{(k)}, p_{\mathcal{F}^\pm}^{(k)}, \vec{b}^{(k)}, \vec{a}, \vec{g}^{(k)})$, $V^{(k)}, V^{\dagger(k)} \in [0, 1]$	$ V^{(k)}\rangle$: <i>+ve valence</i> $ V^{\dagger(k)}\rangle$: <i>−ve valence</i>

S1.2.1 Relevance

Specifically, relevance and irrelevance are evaluated from a set of (1) goal-and-action-based rules and (2) appraisal-variables-based rules. For the former, an action $i^{(k)}$ for feature k may be chosen from a set of choice of action $\mathcal{A}^{(k)}$ where $i^{(k)}$ can be mapped to some action tendency in a set $\mathcal{T}$ via $t : \mathcal{A}^{(k)} \rightarrow \mathcal{T}$ (see *Satisfaction* in Section S1.3.3 for details). On the other hand, a goal j from a set $\mathcal{G}$, *desired* or *undesired*, defines what the agency wants to achieve or avoid. The goal is *important* whether it is desired or undesired. The action $i^{(k)}$ has an effect on (*affects*) goal j if it either *facilitates* or *inhibits* the goal. Two statements concerning a particular feature k can be set up: ζ_1 : ANY $i^{(k)} \in \mathcal{A}^{(k)}$ affects goal j , and ζ_2 : goal j is important. The agency may *agree* or *disagree* to ζ_1 and ζ_2 . *Agree* and *disagree* act as a bidimensional unipolar scale as one can simultaneously agree to the statements in some aspects and

disagree in other aspects. The three statements result in the following rule:

$$\begin{aligned}
& \text{IF} && \left(\zeta_1 : \bigvee_{i^{(k)} \in \mathcal{A}^{(k)}} (\text{action } i^{(k)} \text{ affects goal } j) \right) \\
& && \bigwedge \quad \zeta_2 : \text{goal } j \text{ is important} \\
& && \bigwedge \quad \zeta_3 : \text{agree to } (\zeta_1 \wedge \zeta_2) \\
& \text{THEN} && \text{feature } k \text{ is relevant (under goal } j),
\end{aligned} \tag{S2}$$

or, abbreviated in terms of membership function, $R_j^{(k)} = \mu_R(\vec{b}^{(k)}, a_j, g_j^{(k)})$, where $R_j^{(k)}$ is the value of relevance under goal j and μ_R is the corresponding membership function. For the appraisal-variables-based rules, the agency checks separately the four appraisal variables. If the feature possesses either the indicative or the counter-indicative appraisal property, the feature is considered relevant. Considering all appraisal variables, the rule becomes

$$\begin{aligned}
& \text{IF} && \zeta_4 : \bigvee_{l \in \mathcal{D}} \text{feature } k \text{ is } l \\
& \text{THEN} && \text{feature } k \text{ is relevant (due to encoded features),}
\end{aligned} \tag{S3}$$

or, abbreviated in terms of membership function, $R_p^{(k)} = \mu_R(\vec{p}^{(k)})$, where $R_p^{(k)}$ is the value of relevance due to encoded features. The disjunction of the above 2 rules (Eqs. (S2) and (S3)) for all goals gives the resultant relevance, i.e.,

$$\begin{aligned}
& \text{IF} && \bigvee_{j \in \mathcal{G}} \left(\text{feature } k \text{ is relevant under goal } j \right) \\
& && \bigvee \quad \text{feature } k \text{ is relevant due to encoded features} \\
& \text{THEN} && \text{feature } k \text{ is relevant,}
\end{aligned} \tag{S4}$$

or, abbreviated in terms of membership function, $R^{(k)} = \mu_R(\vec{R}^{(k)}, R_p^{(k)})$, where $R^{(k)}$ is the value of relevance and $\vec{R}^{(k)} = (R_j^{(k)})$. The two sets of rules are similar for irrelevance evaluation. The negation of the conditions leads to irrelevance. Nonetheless, for *agree* and *disagree* as a pair of bidimensional unipolar scales (one can simultaneously agree and disagree to one entity given multiple perspectives), we

consider the negation of *agree* to be *disagree*. Therefore,

$$\begin{aligned}
\text{IF} \quad & \neg \left[\left(\zeta_1 : \bigvee_{i^{(k)} \in \mathcal{A}^{(k)}} (\text{action } i^{(k)} \text{ affects goal } j) \right) \right. \\
& \quad \left. \wedge \left(\zeta_2 : \text{goal } j \text{ is important} \right) \right] \\
& \vee \zeta_3^\dagger : \text{disagree to } (\zeta_1 \wedge \zeta_2) \\
\text{THEN} \quad & \text{feature } k \text{ is irrelevant (under goal } j),
\end{aligned} \tag{S5}$$

or, abbreviated in terms of membership function, $R_j^{\dagger(k)} = \mu_{R^\dagger}(\vec{b}^{(k)}, a_j, g_j^{(k)})$, where $R_j^{\dagger(k)}$ is the value of relevance under goal j and $\mu_{R^\dagger}$ is the corresponding membership function, and

$$\begin{aligned}
\text{IF} \quad & \neg \left(\zeta_4 : \bigvee_{l \in \mathcal{D}} \text{feature } k \text{ is } l \right) \\
\text{THEN} \quad & \text{feature } k \text{ is irrelevant (due to feature encoding)}.
\end{aligned} \tag{S6}$$

Note that the value of irrelevance due to encoded features is $1 - R_p^{(k)}$. The disjunction of the above 2 rules (Eqs. (S5) and (S6)) gives the resultant irrelevance, i.e.,

$$\begin{aligned}
\text{IF} \quad & \bigvee_{j \in \mathcal{G}} \left(\text{feature } k \text{ is irrelevant under goal } j \right) \\
& \vee \text{feature } k \text{ is irrelevant due to encoded features} \\
\text{THEN} \quad & \text{feature } k \text{ is irrelevant,}
\end{aligned} \tag{S7}$$

or, abbreviated in terms of membership function, $R^{\dagger(k)} = \mu_{R^\dagger}(\vec{R}^{\dagger(k)}, R_p^{(k)})$, where $R^{\dagger(k)}$ is the value of irrelevance and $\vec{R}^{\dagger(k)} = (R_j^{\dagger(k)})$.

S1.2.2 Valence

Valence evaluation relies primarily on the available beliefs $(b_{ij}^{(k)})$ of whether an action $i^{(k)}$ facilitates $(b_{ij}^{(k)} > 0)$ or inhibits $(b_{ij}^{(k)} < 0)$ the goal, the ambition to the goal (a_j) (desired $(a_j > 0)$ or undesired $(a_j < 0)$), and the agreement $(g'_{ij}{}^{(k)})$ of the agency to the former two statements (agree $(g'_{ij}{}^{(k)} > 0)$ or disagree $(g'_{ij}{}^{(k)} < 0)$). Depending on the combination of facilitates/inhibits, desired/undesired and

agree/disagree, the positive and negative valence are evaluated as follows:

$$\begin{aligned}
& \text{IF} \quad \zeta'_1 : \text{feature } k \text{ is } \begin{cases} \text{aid} \\ \text{obstacle} \end{cases} \\
& \quad \wedge \quad \zeta'_2 : \text{feature } k \text{ is } \begin{cases} \text{good} \\ \text{bad} \end{cases} \\
& \quad \wedge \quad \zeta'_3 : \text{action } i^{(k)} \begin{cases} \text{facilitates, } \text{sgn}(b_{ij}^{(k)}) = +1 \\ \text{inhibits, } \text{sgn}(b_{ij}^{(k)}) = -1 \end{cases} \text{ goal } j \\
& \quad \wedge \quad \zeta'_4 : \text{goal } j \text{ is } \begin{cases} \text{desired, } \text{sgn}(a_j) = +1 \\ \text{undesired, } \text{sgn}(a_j) = -1 \end{cases} \\
& \quad \wedge \quad \zeta'_5 : \begin{cases} \text{agree, } \text{sgn}(g'_{ij}^{(k)}) = +1 \\ \text{disagree, } \text{sgn}(g'_{ij}^{(k)}) = -1 \end{cases} \text{ to } (\zeta'_1 \wedge \zeta'_2 \wedge \zeta'_3 \wedge \zeta'_4) \\
& \text{THEN} \quad \text{feature } k \text{ has } \begin{cases} \text{positive, } \text{sgn}(b_{ij}^{(k)} a_j g'_{ij}^{(k)}) = +1 \\ \text{negative, } \text{sgn}(b_{ij}^{(k)} a_j g'_{ij}^{(k)}) = -1 \end{cases} \text{ valence (under goal } j \text{ and action } i^{(k)})
\end{aligned} \tag{S8}$$

where $\text{sgn}(x)$ is the sign function. For a feature with 4 possible actions, Eq. (S8) is a condensation of $2 \times 2 \times (4 \times 2) \times 2 \times 2 = 128$ logical statements per goal per feature, containing 64 for each positive and negative valence. The resultant positive or negative valence is the disjunction of the respective statements, i.e.,

$$\begin{aligned}
& \text{IF} \quad \bigvee_{j \in \mathcal{G}} \bigvee_{i^{(k)} \in \mathcal{A}^{(k)}} \left(\text{feature } k \text{ has } \begin{cases} \text{positive} \\ \text{negative} \end{cases} \text{ valence under goal } j \text{ and action } i^{(k)} \right) \\
& \text{THEN} \quad \text{feature } k \text{ has } \begin{cases} \text{positive} \\ \text{negative} \end{cases} \text{ valence,}
\end{aligned} \tag{S9}$$

or, abbreviated in terms of membership function, $V^{(k)} = \mu_V(p_{\mathcal{E}^\pm}^{(k)}, p_{\mathcal{F}^\pm}^{(k)}, \vec{b}^{(k)}, \vec{a}, \vec{g}^{(k)})$, where $V^{(k)}$ is the value of positive valence, and $V^{\dagger(k)} = \mu_{V^\dagger}(p_{\mathcal{E}^\pm}^{(k)}, p_{\mathcal{F}^\pm}^{(k)}, \vec{b}^{(k)}, \vec{a}, \vec{g}^{(k)})$, where $V^{\dagger(k)}$ is the value of negative valence. μ_V and $\mu_{V^\dagger}$ are the corresponding membership functions.

The measures in the encode phase – mediated by relevance and valence in the comparison phase and moderated by *similarity* between the features of robot and user (similar or dissimilar) – determine the robot's responses Fig. 1.^[1] However, empirical research showed that relevance and valence are more important to engagement than similarity.^[1] For the sake of simplicity, we modeled relevance and valence alone.

S1.3 Response

In the response phase, the robot generates rational cognitions about use intentions and runs affective processes for engagement. Cognition and affect are then collectively considered to estimate the expected

satisfaction. Decision making is done based on the expected satisfaction upon various response strategies (positive and negative approach, change, avoid, etc.)

S1.3.1 Use intentions

In the *response* phase, the robot calculates a value for the so-called *use intentions*, the willingness to employ the user or another agency as a means to achieve robot goals. The variables involved in the evaluation of use intentions are summarized in Table S3. *Utility* as an intermediate variable to indicate the usefulness of the feature to the agency with respect to its goals is calculated first. The expected utility is given by

$$u'_{ij}{}^{(k)} = b_{ij}^{(k)} a_j. \quad (\text{S10})$$

The *mean expected utility* over goals for an action $i^{(k)}$ can then be calculated, subject to a weight for each goal $\Theta^{(k)}(u'_{ij}{}^{(k)})$:

$$\bar{u}_i'{}^{(k)} = \frac{\sum_{j \in \mathcal{G}} \Theta^{(k)}(u'_{ij}{}^{(k)}) \cdot u'_{ij}{}^{(k)}}{\sum_{j \in \mathcal{G}} \Theta^{(k)}(u'_{ij}{}^{(k)})}. \quad (\text{S11})$$

Then, the *indicative utility* ($u'_+{}^{(k)}$) and *counter-indicative utility* ($u'_-{}^{(k)}$) values of the feature k are determined by the highest and lowest mean utility for the action $i^{(k)}$:

$$u'_\pm{}^{(k)} = \max_{i^{(k)} \in \mathcal{A}^{(k)}} (\pm \bar{u}_i'{}^{(k)}, 0). \quad (\text{S12})$$

Use intentions are modeled as a complex composition of various factors, each carrying a particular weight.

The indicative ($u'_+{}^{(k)}$) and counter-indicative ($u'_-{}^{(k)}$) use intentions can be evaluated by

$$\vec{u}^{(k)} = \begin{bmatrix} u'_+{}^{(k)} \\ u'_-{}^{(k)} \end{bmatrix} = \mathbf{B}_{\text{ui}} \vec{r}_{\text{ui}}^{(k)} \quad (\text{S13})$$

where the 2×10 matrix $\mathbf{B}_{\text{ui}}$ is the weight matrix for use intentions, containing the weights for $\vec{r}_{\text{ui}}^{(k)}$, the component vector for use intention in $\mathbb{R}^{10}$:

$$\vec{r}_{\text{ui}}^{(k)} = \text{vec} \left(\left\{ \begin{bmatrix} u'_+{}^{(k)} \\ u'_-{}^{(k)} \end{bmatrix}, \begin{bmatrix} u'_+{}^{(k)} \\ u'_-{}^{(k)} \end{bmatrix} \otimes \begin{bmatrix} s & p_{S^+}^{(k)} \\ s^\dagger & p_{S^-}^{(k)} \end{bmatrix} \right\} \right) \quad (\text{S14})$$

where $\otimes$ denotes the Kronecker product and s and $s^\dagger$ are the predefined similarity and dissimilarity. Abuse of notation is used to represent the construction of $\vec{r}_{ui}^{(k)}$ as the vectorization of the two matrices in a row, from top to bottom and from left to right.

Table S3: Variables involved in the evaluation of use intentions in the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system.

Name	Si-Coppélia notation	Description	Qubit notation
Expected utility	$u'_{ij}{}^{(k)}$	$u'_{ij}{}^{(k)} \in [-1, 1]$	$ u'_{ij,+}{}^{(k)}\rangle$: <i>indicative</i> $ u'_{ij,-}{}^{(k)}\rangle$: <i>counter-indicative</i>
Mean expected utility over goals	$\bar{u}'_i{}^{(k)}$	$\bar{u}'_i{}^{(k)} \in [-1, 1]$	$ \bar{u}'_{i,+}{}^{(k)}\rangle$: <i>indicative</i> $ \bar{u}'_{i,-}{}^{(k)}\rangle$: <i>counter-indicative</i>
Indicative & Counter-indicative utility	$u'_+{}^{(k)}, u'_-{}^{(k)}$	$u'_+{}^{(k)}, u'_-{}^{(k)} \in [0, 1]$	$ u'_+{}^{(k)}\rangle$: <i>indicative</i> $ u'_-{}^{(k)}\rangle$: <i>counter-indicative</i>
Weight function for $u'_{ij}{}^{(k)}$	$\Theta_i^{(k)}(u'_{ij}{}^{(k)})$	$\sum_j \Theta_j^{(k)}(u'_{ij}{}^{(k)}) = 1$	Embedded in R_y rotations
Component vector for use intentions	$\vec{r}_{ui}^{(k)}$	10 components	Formed by constitute qubits as controls
Weight matrix for use intentions	$\mathbf{B}_{ui}$	2×10 matrix, $\sum_i (\mathbf{B}_{ui})_{ij} \leq 1$	Embedded in R_y rotations
Use intentions (UI)	$\vec{u}^{(k)} = (u_+^{(k)}, u_-^{(k)})^T$	$u_+^{(k)}, u_-^{(k)} \in [0, 1]$	$ u_+^{(k)}\rangle$: +ve UI $ u_-^{(k)}\rangle$: -ve UI

S1.3.2 Engagement as involvement and distance

The robot establishes engagement with an agency by calculating the levels of *involvement* with and *distance* towards, for instance, its user. Involvement and distance are two tendencies that occur in parallel and may compensate one another. This trade-off in the engagement process is accompanied by specific emotions, which may differ in each particular case. The variables involved in the comparison

phase are summarized in Table S4. Similar to use intentions, involvement–distance is modeled as a complex composition of various factors, each carrying a particular weight. The involvement ($E_{\text{inv}}^{(k)}$) and distance ($E_{\text{dist}}^{(k)}$) is given by

$$\vec{E}^{(k)} = \begin{bmatrix} E_{\text{inv}}^{(k)} \\ E_{\text{dist}}^{(k)} \end{bmatrix} = \mathbf{B}_{\text{id}} \vec{r}_{\text{id}}^{(k)} \quad (\text{S15})$$

where the 2×26 matrix $\mathbf{B}_{\text{id}}$ is the weight matrix for involvement–distance, containing the weights for $\vec{r}_{\text{id}}^{(k)}$, the component vector in $\mathbb{R}^{26}$:

$$\vec{r}_{\text{id}}^{(k)} = \text{vec} \left(\left\{ \begin{bmatrix} p_{\mathcal{S}^+}^{(k)} & p_{\mathcal{P}^+}^{(k)} & p_{\mathcal{F}^+}^{(k)} & R^{(k)} & V^{(k)} \\ p_{\mathcal{S}^-}^{(k)} & p_{\mathcal{P}^-}^{(k)} & p_{\mathcal{F}^-}^{(k)} & R^{\dagger(k)} & V^{\dagger(k)} \end{bmatrix}, \begin{bmatrix} R^{(k)} & V^{(k)} \\ R^{\dagger(k)} & V^{\dagger(k)} \end{bmatrix} \otimes \begin{bmatrix} p_{\mathcal{S}^+}^{(k)} & s \\ p_{\mathcal{S}^-}^{(k)} & s^{\dagger} \end{bmatrix} \right\} \right) \quad (\text{S16})$$

where the convention of vectorization is the same as that in Eq. (S13). The involvement–distance trade-off parameter is then calculated as weighed mean of the *fuzzy-or* operator and the compensation between involvement with distance and the compensation factor for the two (β_{idt}):

$$E'_{\text{idt}}^{(k)} = \beta_{\text{idt}} \max(E_{\text{inv}}^{(k)}, E_{\text{dist}}^{(k)}) + (1 - \beta_{\text{idt}}) \frac{E_{\text{inv}}^{(k)} + E_{\text{dist}}^{(k)}}{2} \quad (\text{S17})$$

Table S4: Variables in the involvement–distance trade-off of the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system.

Name	Si-Coppélia notation	Description	Qubit notation
Component vector for involvement–distance	$\vec{r}_{\text{id}}^{(k)}$	26 components	Formed by constitute qubits as controls
Weight matrix for involvement–distance	$\mathbf{B}_{\text{id}}$	2×26 matrix, $\sum_i (\mathbf{B}_{\text{id}})_{ij} \leq 1$	Embedded in R_y rotations
Involvement–distance	$\vec{E}^{(k)} = (E_{\text{inv}}^{(k)}, E_{\text{dist}}^{(k)})^T$	$E_{\text{inv}}^{(k)}, E_{\text{dist}}^{(k)} \in [0, 1]$	$ E_{\text{inv}}^{(k)}\rangle$: <i>involvement</i> $ E_{\text{dist}}^{(k)}\rangle$: <i>distance</i>

Table S4: Variables in the involvement–distance trade-off of the Silicon Coppélia system, and their corresponding kets in the Quantum Coppélia system. (Cont’d)

Name	Si-Coppélia notation	Description	Qubit notation
Compensation factor for involvement & distance	β_{idt}	$\beta_{\text{idt}} \in [0, 1]$	$ \beta_{\text{idt}}\rangle$
Involvement– distance trade-off	$E'_{\text{idt}}^{(k)}$	$E'_{\text{idt}}^{(k)} \in [0, 1]$	$ E'_{\text{idt}}^{(k)}\rangle$

S1.3.3 Satisfaction

Use intentions, involvement, and distance together determine the overall *satisfaction* of the robot with its user or another agency. The level of satisfaction determines the robot’s decision of whether and how it is going to interact with the user (or avoid – stop and turn to another). The variables involved in the calculation are summarized in Table S5. The decision-making process is carried out by calculating the *expected satisfaction* of interacting (again) with a user, which is the weighed mean of the involvement–distance trade-off (IDT) and the average of the indicative and counter-indicative use intentions with the respective weights σ'_{idt} and σ'_{ui} :

$$S^{(k)} = \sigma'_{\text{idt}} E'_{\text{idt}}^{(k)} + \sigma'_{\text{ui}} \frac{u_+^{(k)} + u_-^{(k)}}{2}. \quad (\text{S18})$$

Eq. (S18) reveals that the robot makes a decision on the more rationally generated use intentions in unison with the more affectively generated involvement–distance trade-off. The feature k_{max} that promises the highest expected satisfaction during the interaction is selected, i.e., treating $S^{(k)}$ as a function of k ,

$$k_{\text{max}} = \arg\left(\max_{k \in \mathcal{F}}(S^{(k)})\right). \quad (\text{S19})$$

After selecting the feature of the highest satisfaction, the robot inspects the overt actions $i^{(k_{\text{max}})}$ available to this feature, and chooses an action tendency (as noted in *relevance* and *valence* in Section S1.2.1)

from the following 4 types:

1. positive approach (e.g., compliment the user): $t(i) = p$
2. negative approach (e.g., criticize the user): $t(i) = n$
3. change (e.g., teach the user): $t(i) = c$
4. avoid the user: $t(i) = a$.

The 4 types of action tendencies are depicted in the report of Silicon Coppélia.^[2] In general, a feature may not offer all the 4 action tendencies. Additional action tendencies are also possible (e.g., do nothing).

The selection of the action tendency is done by calculating the respective expected satisfaction upon the action:

$$S_i^{(k_{\max})} = \begin{cases} \sigma_{i,p} E_{\text{inv}}^{(k_{\max})} + \sigma_{d,p} (1 - E_{\text{dist}}^{(k_{\max})}) + \sigma_{u,p} \bar{u}'_i^{(k_{\max})}, & t(i) = p \\ \sigma_{i,n} (1 - E_{\text{inv}}^{(k_{\max})}) + \sigma_{d,n} E_{\text{dist}}^{(k_{\max})} + \sigma_{u,n} \bar{u}'_i^{(k_{\max})}, & t(i) = n \\ \sigma_{i,c} E_{\text{inv}}^{(k_{\max})} + \sigma_{d,c} E_{\text{dist}}^{(k_{\max})} + \sigma_{u,c} \bar{u}'_i^{(k_{\max})}, & t(i) = c \\ \sigma_{i,a} (1 - E_{\text{inv}}^{(k_{\max})}) + \sigma_{d,a} E_{\text{dist}}^{(k_{\max})} + \sigma_{u,a} \bar{u}'_i^{(k_{\max})}, & t(i) = a \end{cases} \quad (\text{S20})$$

or, in matrix representation,

$$\begin{bmatrix} S_{i|t(i)=p}^{(k_{\max})} \\ S_{i|t(i)=n}^{(k_{\max})} \\ S_{i|t(i)=c}^{(k_{\max})} \\ S_{i|t(i)=a}^{(k_{\max})} \end{bmatrix} = \begin{bmatrix} \sigma_{i,p} & -\sigma_{d,p} & \sigma_{u,p} & \sigma_{d,p} \\ -\sigma_{i,n} & \sigma_{d,n} & \sigma_{u,n} & \sigma_{i,n} \\ \sigma_{i,c} & \sigma_{d,c} & \sigma_{u,c} & 0 \\ -\sigma_{i,a} & \sigma_{d,a} & \sigma_{u,a} & \sigma_{i,a} \end{bmatrix} \begin{bmatrix} E_{\text{inv}}^{(k_{\max})} \\ E_{\text{dist}}^{(k_{\max})} \\ \bar{u}'_i^{(k_{\max})} \\ 1 \end{bmatrix}. \quad (\text{S21})$$

The action with the highest expected satisfaction will be chosen as the final response, i.e., treating $S_i^{(k_{\max})}$ as a function of $i^{(k_{\max})}$,

$$i_{\max} = \arg \left(\max_{i \in \mathcal{A}^{(k_{\max})}} (S_i^{(k_{\max})}) \right). \quad (\text{S22})$$

Table S5: Variables in the satisfaction evaluation of the Silicon Coppélia system, and their corresponding kets in Quantum Coppélia.

Name	Si-Coppélia notation	Description	Qubit notation
Weight factors for expected satisfaction	$\sigma'_{\text{idt}}, \sigma'_{\text{ui}}$	$\sigma'_{\text{idt}}, \sigma'_{\text{ui}} \in [0, 1]$	Embedded in R_y rotations
Expected satisfaction	$S^{(k)}$	$S^{(k)} \in [0, 1]$	$ S^{(k)}\rangle$
Feature contributing to the greatest expected satisfaction	k_{max}	$k_{\text{max}} \in \mathcal{F}$	$ k_{\text{max}}\rangle$
Component weight for expected satisfaction for action $i^{(k)}$	$\sigma_{\mu,\nu}$	$\mu = \text{i, d, u}; \nu = \text{p, n, c, a}$ $\sum_{\mu} \sigma_{\mu,\nu} \leq 1$	Embedded in R_y rotations
Expected satisfaction for action $i^{(k)}$	$S_i^{(k)}$	$S_i^{(k)} \in [-1, 1]$	$ S_i^{(k)}\rangle$
Action contributing to the greatest expected satisfaction $S_i^{(k_{\text{max}})}$	i_{max}	$i_{\text{max}} \in \mathcal{A}^{(k_{\text{max}})}$	$ i_{\text{max}}\rangle$

References

- [1] Van Vugt, H. C., Hoorn, J. F. & Konijn, E. A. Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds* **20**, 195–204; <https://doi.org/10.1002/cav.312> (2009).
- [2] Hoorn, J. F., Baier, T., Van Maanen, J. A. N. & Wester, J. Silicon Coppélia and the formalization of the affective process. *IEEE Transactions on Affective Computing*, 1; <https://doi.org/10.1109/TAFCC.2020.3048587> (2021).