Early Humorous Interaction: Towards a Formal Model

Yingqin Hu[1,2], Jonathan Ginzburg[1,2], Catherine Pelachaud[1,3]

[1] Centre national de la recherche scientifique (CNRS)
[2] Laboratoire Linguistique Formelle (LLF), Université Paris-Cité
[3] Institut des Systèmes Intelligents et de Robotiques (ISIR), Sorbonne Université
yingqinhu@gmail.com, yonatan.ginzburg@u-paris.fr, catherine.pelachaud@sorbonne-universite.fr

Abstract

Current computational models for humour recognition and laughter generation in dialogue systems face significant limitations in explainability, context consideration and adaptability. This paper approaches these challenges by investigating how humour recognition develops in its earliest forms—during the first year of life. Drawing on developmental psychology and cognitive science, we propose a formal model incorporated within the KoS dialogue framework. This model captures how infants evaluate potential humour through knowledge-based appraisal and context-dependent modulation, including safety, emotional state, and social cues. Our model formalises dynamic knowledge updates during the dyadic interaction. We believe that this formal model can serve as the basis for developing more natural humour appreciation capabilities in dialogue systems and can be implemented in a robotic platform.

1 Introduction

Laughter serves as a critical non-verbal component in human interaction, signalling recognition of incongruity and creating social affiliation (Ginzburg et al., 2020; Mazzocconi et al., 2020). Integrating laughter capabilities into dialogue systems remains challenging due to the subjective and context-dependent nature of humour. Current approaches to laughter generation predominantly rely on neural network models, including reinforcement learning (Weber et al., 2018) and Long Short-Term Memory networks (Sun et al., 2018), which generate responses based on latent values but often lack interpretability.

Recent work has examined temporal sequence patterns of laughter in dialogues (El Haddad et al., 2019), developed virtual conversational agents that predict appropriate moments for smiling or laughing (Ritschel et al., 2020), and proposed attention mechanisms that analyse wording context to place contextually appropriate laughter (Xu et al., 2021). However, these approaches face significant limitations: 1) **Limited explainability**: neural approaches often function as black boxes, making it difficult to understand the precise reasoning behind laughter generation; 2) **Inadequate consideration of context**: many models fail to account for the rich multimodal social signals, safety factors and mood that inform humour interpretation, and 3) **Poor cognitive adaptability**: current systems struggle to adjust to shifting conversational contexts and participant relationships.

We propose a formal model based on a principled, mechanistic-approach to explicate the development of humour recognition and laughter production. We use an example extracted from the Rollins corpus to demonstrate the model's capacity to explain the emergence of specific emotional responses in infants towards each caregiver-produced event. By focusing on the first year of life, when infants begin to perceive and respond to incongruity, we can identify the core cognitive mechanisms that underlie humour appreciation before they become overly complex or culturally specific.

The paper is organised as follows. First, in Section 2, we review foundational theories of humour development in infants and a potential framework to model this humour interaction. Next, we propose our research questions in Section 3 and describe our methodological approach in Section 4. We then present our formal model in Section 5, detailing both its architecture and appraisal process, followed by a comprehensive case study in Section 6. Finally, we conclude our work and its contributions (Section 7), and discuss its limitations and directions for future research (Section 8).

Proceedings of the 26th Annual Meeting of SIGDIAL, pages 254–267
Aug 25–27, 2025.

2 Background: humour-based laughables in the first year

2.1 Humour from common-sense knowledge

Appreciating humour is cognitive in nature, hence requires the detection of at least one clash between what is perceived and what is considered as normal – *incongruity* (Raskin, 1979; McGhee and Pistolesi, 1979; Shultz, 2017).

Longitudinal studies show that humour during the first year mainly stems from violations of four types of infant common-sense knowledge (the norm): *Human* (expectations about human behaviour), *Object* (physical properties and behaviours of objects), *Social interaction* (norms and expectations in social contexts), and *Metalinguistic* (expectations about language and communication patterns) (Sroufe and Wunsch, 1972; Horgan, 1981; Reddy, 1991; Hoicka and Gattis, 2008; Mireault et al., 2012; Hoicka et al., 2022; Hu et al., 2024).

2.2 The degree of discrepancy and its influence towards pleasure derived

However, merely detecting a clash is insufficient for deriving pleasure; the discrepancy must be within an appropriate range – neither too difficult nor too easy relative to the child's current knowledge state, suggested by Vygotsky and Cole (1978)'s zone of proximity, Suls (1972)'s incongruity-resolution (IR) model, and the arousal theory of laughter (Spencer, 1875; Berlyne, 1960; Rothbart, 1973). Empirical studies (Shultz, 1972, 1974; McGhee, 1976) further demonstrate that the pleasure derived from such discrepancy follows an inverted U-shaped pattern – the pleasure peaks when the degree of discrepancy is at an optimal discrepancy value.

2.3 Factors influencing humour appreciation

Apart from the infant's self-knowledge, humour appreciation is contextual in nature, influenced by environmental and interlocutor safety, current emotional state, and social cues.

Environmental and interlocutor safety Safety represents a prerequisite for positive appraisals. Research consistently demonstrates that infants must perceive their environment as safe before they can engage in humour appreciation (Bronson, 1972; Sroufe et al., 1974; Baillargeon et al., 1985; Mireault and Reddy, 2020). Without this foundational sense of safety, even potentially humorous stimuli may be appraised as threatening rather than enjoyable.

Emotional state An infant's current affective state serves as a contextual filter through which all stimuli are evaluated. Infants experiencing negative emotional states may find it challenging to appreciate certain incongruities that might otherwise elicit positive responses in more neutral or positive emotional contexts (Sullivan and Lewis, 1989; Legerstee, 1997; Braarud and Stormark, 2006; Soussignan et al., 2006).

Social cues The evaluation will also influenced by caregiver cues. From approximately 5.5 months, infants use dyadic caregiver interactions, especially emotional signalling and contingent responsiveness, to interpret perceived events (Sorce et al., 1985; Mumme et al., 1996), including exaggerated pitch contours, increased amplitude variability (Hoicka and Gattis, 2012), and the parents' smile or laugh (Mireault et al., 2014, 2015).

2.4 Humour from episodic knowledge

Beyond common-sense knowledge, infants develop episodic knowledge through repeated exposure to sequential events, enabling them to transform humorous stimuli into local laughables—through prediction or by generating novel humour.

Prediction One-month-old infants already demonstrate a capacity to predict event outcomes (Köster et al., 2020) and sensitivity to the temporal sequencing of the visual events (Canfield and Haith, 1991). These abilities improve significantly from 8 months onward, as shown by infants tracking transitional probabilities between syllables (Saffran et al., 1996).

Teasing Reddy 1 (2001); Reddy and Mireault (2015) demonstrate that infants function as active creators of humour, not merely perceivers. Longitudinal studies with infants aged 7-11 months reveal that 87% of infants at both 8 and 11 months exhibited clear "clowning" behaviours, intentionally repeating actions to re-elicit laughter from others.

3 Research questions

Given the interplay of cognitive, affective and social factors in infant humour interaction presented on our literature review, we hypothesise that humour interactions in infants can be modelled using the architecture of the OCC model (Ortony et al., 2022), widely used to model emotions in artificial intelligence and which has significantly influenced computational models of emotion (e.g., (Gebhard, 2005; Marsella and Gratch, 2009; El-Nasr et al., 2000)). This model emphasises that the emotional responses arise from cognitive appraisals of events, agents, and objects based on the agent's beliefs or relationship with its environment (*appraisal derivation* $\rightarrow$ *affect derivation*). Corresponding cognitive and/or behavioural changes occur alongside the derived affect *affect consequence*, and these changes influence the next appraisal process, forming a loop. This framework, therefore, can serve as a basic approach to model an infant's dynamic humour appreciation – the transition from experiencing humour events with cognitive challenge to appreciating them and generating emotional response arises from local humour laughable (predicting humour events or applying updated knowledge to create novel humour).

However, as the OCC model covers all emotions and is also not specific to infants, additional refinements are needed. More detailed components require specification, including: 1) how to represent infant knowledge and detect conflicts between this knowledge and what they perceive; 2) how infants combine self-generated appraisals with contextual factors; and 3) how different emotional responses to humour events generate corresponding adaptations in knowledge and behaviour.

4 Method

Based on our hypothesis, we propose an OCC-inspired model of infant-caregiver humour interaction using the KoS[1] dialogue framework (Larsson, 2002; Ginzburg and Cooper, 2004; Purver, 2006; Fernández Rovira, 2006; Ginzburg and Fernández, 2010; Ginzburg, 2012, 2016) (see Appendix A.3 for details) which is formulated within Type Theory with Records (TTR) (Cooper, 2005, 2012; Cooper and Ginzburg, 2015; Cooper, 2023) (see Appendix A.1 for details). And we exemplify our model by analysing one long-sequence of a play interaction extracted from the Rollins corpora (Rollins, 2003).

[1]KoS is not an acronym, though related to conversationally oriented semantics.

5 The model

5.1 Architecture

Based on the literature review (Section 2), an infant's appreciation of humour and the intensity of pleasure derived from this seems to relate to the degree of discrepancy between the humorous event and their self-knowledge. This discrepancy value is then modulated by three contextual factors: environmental and interlocutor safety, the infant's current emotional state, and social cues from the interlocutor (*appraisal derivation*). The corresponding emotional response and pleasure are derived from this modulated discrepancy value (*affect derivation*), followed by a knowledge and behaviour update that influences the next appraisal process (*affect consequence*).

To effectively model humorous interactions between caregiver and infant, our framework requires: 1) a state that records the externalized information that one agent (caregiver/infant) perceives *(input)* and produces (*output*), and 2) a state that records one agent's current knowledge and mood and how perceived events influence these two (*process*).

We therefore implement this conceptualization by adapting the KoS dialogue framework, which proposes using two dual states (*public* and *private*) for interlocutors, where *public* relates to externalized information and *private* what is not externalized. These two states are combined to form a *TotalInformationState* (TIS) for each interactional agent (caregiver and infant). The two *TotalInformationState* (TIS) states are then composed to form an *InteractionState* (IS) for the dyadic interaction. Figure 1 illustrates the architecture and update flow of our model.

$$\text{IS} =_{def} \begin{bmatrix} \text{caregiver} : \begin{bmatrix} \text{TIS} \begin{bmatrix} \text{private} : \text{PrivateType} \\ \text{public} : \text{PublicType} \end{bmatrix} \end{bmatrix} \\ \text{infant} : \begin{bmatrix} \text{TIS} \begin{bmatrix} \text{private} : \text{PrivateType} \\ \text{public} : \text{PublicType} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The private state includes *knowledge*[2] (eq. 1) and *mood*(serving as the emotional state)[3]. For *knowledge* state, we distinguish between defeasible common sense, and episodic knowledge, where

[2]The episodic memory for a child is by default short-term, as children in this stage haven't fully developed long-term memory capacity (Rubin, 2000; Ribordy et al., 2013).

[3]Following Russell (2003), *mood* is treated as a weighted sum of three dimensions (*pleasure*, *responsible* and *power*) with each value in N in (Ginzburg et al., 2020). As this model is for the infant during the first year, and to facilitate computation, we left only *pleasure* and *control* and set these two values range over $[-1, 1]$ in our model.

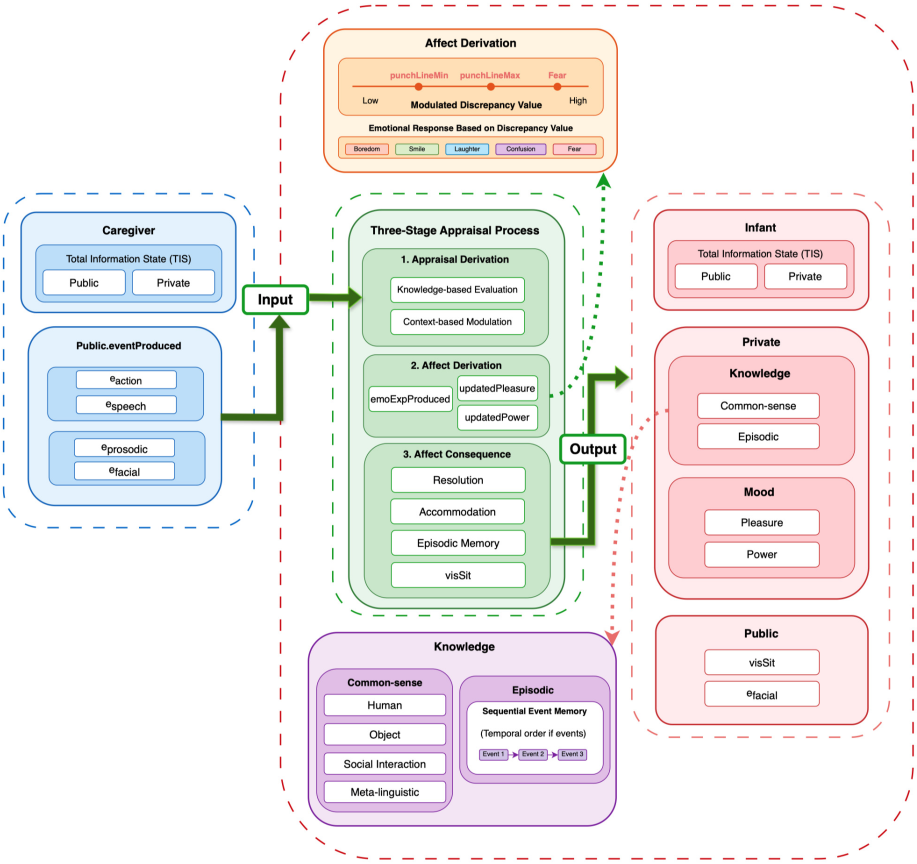


Figure 1: The architecture and update flow of our computational model for infant humour appreciation is illustrated above. With a mutual-attention engagement, the caregiver produces an event e_t to the infant at time t, the infant, as the addressee, takes this event e_t as **input** and then **undergoes a three-stage appraisal process** (green box). First, during the *appraisal derivation* stage (part of the green box), the infant evaluates event e_t using their self-knowledge, followed by modulation of this initial appraisal through contextual factors, resulting in a final discrepancy value ($\hat{\Delta}T_e^t$). Second, during the *affect derivation* stage (part of the green box, detailed in the orange box), the corresponding emotional response (*emoExpProduced*) and mood (*updatedPleasure* and *updatedPower*) are computed based on this modulated discrepancy value ($\hat{\Delta}T_e^t$). Third, the *affect consequence* stage (part of the green box) updates the infant's knowledge ($K_{common\text{-}sense}$ and $K_{episodic}$) (part of the green box, detailed in the purple box) and visual situation (behaviour) (part of the red box Public) (*visSit*, indicating willingness to maintain focus in the interaction). The resulting updated *Interaction State* (IS) is then **output**, preparing for the next turn of the interaction.

common sense knowledge ($K_{common\text{-}sense}$) represents knowledge with little doubt (Bressler and Menon, 2010) used to detect humour, and episodic memory ($K_{episodic}$) records sequence events the agent has been exposed to and can use to create local laughables (predicting humour or teasing others).

Both types of knowledge are represented by collections of topoi τ with associated belief strengths (w_{bel}), while the actual event is defined as the enthymemes (ϵ) (Breitholtz and Cooper, 2011; Breitholtz, 2014, 2020).

$$K_{common-sense} = \{(\tau_i, \omega_i) \mid i = 1, 2, \ldots, n\}$$
$$K_{episodic} = \{(\tau_i, \omega_i) \mid i = 1, 2, \ldots, n\} \quad (1)$$

where ω_i represents the strength of belief in topoi/knowledge τ_i.

The *public* state contains shared temporal information (*Time*), the visual situation (*visSit*), and the event produced (*eventProduced*) by this agent.

$$\text{TIS} =_{def} \begin{bmatrix} \text{private} : \begin{bmatrix} \text{knowledge} & \begin{bmatrix} \text{common sense} : K_{common\text{-}sense} \\ \text{episodic} : K_{episodic} \end{bmatrix} \\ \text{mood} & \begin{bmatrix} \text{pleasure} : [\text{-}1,1] \\ \text{control} : [\text{-}1,1] \end{bmatrix} \end{bmatrix} \\ \text{public} : \begin{bmatrix} \text{time} : \text{Time} \\ \text{visSit} : \text{Ind} \\ \text{eventProduced} : \text{Event} \end{bmatrix} \end{bmatrix}$$

The input/produced event is defined according to what Lücking and Ginzburg (2023) proposed in order to capture multimodal information:[4]

$$\text{Event} =_{def} \begin{bmatrix} spkr & : \text{Ind} \\ addr & : \text{Ind} \\ e_{\text{action}} & : \text{Trajectory} \\ e_{\text{speech}} & : \text{Phone} \\ e_{\text{prosodic}} & : \text{Contour} \\ e_{\text{face}} & : \text{FaceExp} \end{bmatrix}$$

5.2 Appraisal process

Building upon the OCC model (Ortony et al., 2022), we structure the appraisal process in three parts: *appraisal derivation* (5.2.1), *affect derivation* (5.2.2), and *affect consequence* (5.2.3).

Let e_t be an observed event produced by the caregiver at time t (represented as ϵ_e^t), τ_x^{t-1} (before updating) be the violated knowledge with the belief strength ω_x^{t-1}.

5.2.1 Appraisal derivation

The infant's appraisal follows two stages: 1) evaluation based on existing knowledge (common-sense and episodic) and 2) incorporation of contextual factors to derive the final appraisal $\hat{\Delta} T_e^t$.

Knowledge-based evaluation (ΔT_e^t) The initial evaluation based on self-knowledge can be processed as follows. ΔT_e^t represent the degree of discrepancy between the observed event e and the infant's related knowledge τ_x^{t-1}. The degree of discrepancy based on the infant's knowledge is computed as:

$$\Delta T_e^t = 1 - P(\epsilon_e^t \mid \tau_x^{t-1}) \tag{2}$$

Where $P(\epsilon_e^t \mid \tau_x^{t-1})$ is the conditional probability of the observed event given the established knowledge. $\Delta T_e^t \in [0, 1]$ measures the degree of discrepancy. A higher value of ΔT_e^t (approaching 1) represents a significant clash between the perceived event and the infant's knowledge, while a lower value of ΔT_e^t (approaching 0) represents only a slight clash between the perceived event and their knowledge.

[4]To address the need of capturing the social cues from prosodic features and facial expressions, we add one component e_{prosodic} (e_{face} has already been considered).

Context-based modulation This initial discrepancy assessment is then modulated by three contextual factors:

1. Social cues ($c^t_{facialExp}$ and $c^t_{pitchCont}$) Two primary social signals are considered in this model: facial expressions ($c^t_{facialExp}$) and pitch contours ($c^t_{pitchCont}$).

Facial expressions ($c^t_{facialExp}$) convey emotional valence through visual channels, while pitch contours ($c^t_{pitchCont}$) provide prosodic information about the caregiver's communicative intent. Caregivers' positive facial expressions or/and rising pitch contours provide positive signals for event appraisal and therefore decrease the degree of discrepancy, while their negative facial expressions or/and falling pitch contours amplify the perception of expectation violations.

$$\begin{aligned} c^t_{facialExp} &= \text{valence of facial expression at time t} \\ c^t_{pitchCont} &= \text{slope of F0 contour at time t} \end{aligned} \tag{3}$$

Where $c^t_{facialExp}, c^t_{pitchCont} \in [-1, 1]$

2. Safety factors (c_{env} and c_{people}) The safety context influences how infants process potentially threatening or novel incongruities. As the environment and interaction partners remain constant during a given interaction episode, these safety values are static parameters (i.e., they do not change during the interaction).

Safety is determined by the familiarity of people (c_{people}) interacting with the infant and the environment (c_{env}). When infants are situated in familiar environments with known caregivers (high safety context), they demonstrate enhanced capacity to tolerate and integrate higher degrees of violation in their experiences, facilitating more flexible learning and adaptation. Conversely, in unfamiliar environments or with unknown individuals (low safety context), infants exhibit a significantly reduced threshold for appreciating incongruities.

$$\begin{aligned} c_{env} &= \text{degree of environment familiarity} \\ c_{people} &= \text{degree of people familiarity} \end{aligned} \tag{4}$$

Where $c_{env}, c_{people} \in [-1, 1]$

3. Mood (c^{t-1}_{mood}) The infant's current mood intensity (before updated)(c^{t-1}_{mood}) also serves as a modulating factor in the appraisal process. A positive mood may mitigate the negative influence that a high degree of violation brings, while a negative mood may aggravate the negative influence that a high degree of violation brings. The mood is computed as the average of the pleasure and power dimensions of the infant's private emotional state:

$$c_{mood}^{t-1} = \frac{mood.pleasure_{t-1} + mood.power_{t-1}}{2} \quad (5)$$

Where:
$mood.pleasure_{t-1} \in [-1, 1]$ represents the valence dimension (negative to positive) and $mood.power_{t-1} \in [-1, 1]$ represents the sense of control or agency.

Integrated appraisal result ($\hat{\Delta T}_e^t$) The final appraisal result ($\hat{\Delta T}_e^t$) integrates both knowledge-based evaluation and contextual modulation parameters. The degree of influence of each contextual factor on the final result depends on their respective weight:

$$\begin{aligned} \hat{\Delta T}_e^t &= f_{\text{appraisal}}(\Delta T_e^t, \mathbf{c}, \mathbf{w}) \\ &= \Delta T_e^t \cdot \sigma\left(s \cdot \left(1 - \sum_{i \in \mathcal{C}} w_i \cdot c_i\right)\right) \end{aligned} \quad (6)$$

Where:

$$\mathcal{C} = \{c_{facialExp}^t, c_{pitchCont}^t, c_{env}, c_{people}, c_{mood}^t\}$$
$$\sum_i w_i = 1, \quad c_i \in [-1, 1],$$
$$0 \leq \Delta T_e^t \leq 1, \quad s > 0$$

with s representing a scaling factor and σ denoting the sigmoid function that ensures appropriate scaling of the modulation effect.

5.2.2 Affect derivation ($f_{emoExpProduced}$) and emotional state updating ($f_{updatePleasure}$ and $f_{updatePower}$)

Based on three theoretical frameworks—the zone of proximity theory (Vygotsky and Cole, 1978), the arousal theory of laughter (Spencer, 1875; Berlyne, 1960; Rothbart, 1973), and Incongruity-Resolution (IR) theory (Freud, 1960; Suls, 1972)—as well as empirical work (Shultz, 1972, 1974; McGhee, 1976) we postulate five emotional response cases (see Table 1 (the first column)).

The infant produces an emotional response ($emoExpProduced_t = f_{emoExpProduced}(\hat{\Delta T}_e^t)$) towards event e at time t based on an affect derivation function ($f_{emoExpProduced}$) defined as:

$$f_{emoExpProduced}(\hat{\Delta T}_e^t) = \begin{cases} \text{Fear} & \text{if } \hat{\Delta T}_e^t \geq \theta_{fearLine} \\ \text{Confusion} & \text{if } \theta_{punchlineMax} < \hat{\Delta T}_e^t < \theta_{fearLine} \\ \text{Laughter} & \text{if } \theta_{punchlineMin} \leq \hat{\Delta T}_e^t \leq \theta_{punchlineMax} \\ \text{Smile} & \text{if } \hat{\Delta T}_e^t \approx \theta_{punchlineMin} \\ \text{Boredom} & \text{if } \hat{\Delta T}_e^t < \theta_{punchlineMin} \end{cases} \quad (7)$$

The pleasure update function $f_{\text{updatePleasure}}$ is based on McGhee (1976) and McCall and McGhee (1977)'s discrepancy hypothesis, which states that pleasure derived from discrepancy follows an inverted U-shaped pattern—pleasure peaks when the degree of discrepancy is at an optimal value.

$$\begin{aligned} pleasure_t &= f_{\text{updatePleasure}}(pleasure_{t-1}, \hat{\Delta T}_e^t) \\ &= \max(-1, \min(1, pleasure_{\text{new}})) \end{aligned} \quad (8)$$

Where:

$$pleasure_{\text{new}} = \begin{cases} pleasure_{t-1} + \Phi(\hat{\Delta T}_e^t) & \text{if } emoExpProduced_t = \text{Laughter/Smile} \\ pleasure_{t-1} - \Psi(\hat{\Delta T}_e^t) & \text{if } emoExpProduced_t = \text{Fear} \\ pleasure_{t-1} & \text{Otherwise} \end{cases}$$

With proximity function:

$$\theta_{\text{optimal-discrepancy}} = \frac{|\theta_{punchlineMax} - \theta_{punchlineMin}|}{2} + \theta_{punchlineMin}$$
$$\Phi(\hat{\Delta T}_e^t) = |\hat{\Delta T}_e^t - \theta_{\text{optimal-discrepancy}}|$$
$$\Psi(\hat{\Delta T}_e^t) = |\hat{\Delta T}_e^t - \theta_{fearLine}|$$

The power ($power_t$) is updated based on the infant's perceived control over the appraised event e (i.e., the degree to which the discrepancy exceeds the fear threshold (uncontrolled)):

$$\begin{aligned} power_t &= f_{\text{updatePower}}(power_{t-1}, \hat{\Delta T}_e^t) \\ &= power_{t-1} - (\hat{\Delta T}_e^t - \theta_{fearlLine}) \end{aligned} \quad (9)$$

5.2.3 Affect consequence

During this stage, an infant's two types of knowledge (common-sense and episodic) and behaviour (gaze) are updated based on the derived emotional response. (see Table 1 (the third column) in summary)

1. Common-sense knowledge The updating of common-sense knowledge is grounded in Suls (1972)'s Incongruity-Resolution (IR) theory and Piaget et al. (1952)'s schema theory. This updating process relates to both the newly observed knowledge and the old knowledge that is violated.

Given an observed event e at time t and the degree of discrepancy of this event being $\hat{\Delta}T_e^t$:

Resolution (f_{res}) If the conflict remains unresolved (confusion case), the infant, based on observation and existing self-knowledge, attempts to identify a rule that explains the conflict and reduces the discrepancy with the resolved value θ_{res}^t, on which is based the infant's current knowledge and the observation.

$$\begin{aligned} \Delta_e^{t+1} &= f_{res}(\hat{\Delta}T_e^t, \theta_{res}^t) \\ &= \hat{\Delta}T_e^t - \theta_{res}^t \end{aligned} \tag{10}$$

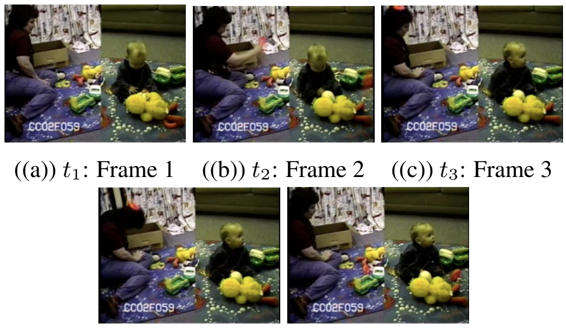
Accommodation (f_{acc}) Conversely, if this conflict is resolved (smile/laughter case), this new knowledge will be accommodated in the common-sense knowledge $K_{common-sense}^{t-1}$ through two processes: 1) Adding this knowledge with a strength of belief (ω_{acc}^t) and 2) reducing the old conflicted knowledge's strength of belief by the same value. The strength of belief depends on whether this knowledge is being accommodated before:

$$\begin{aligned} &K_{common-sense}^t \\ &= f_{acc}(K_{common-sense}^{t-1}, \{(\tau_x^{t-1}, \omega_x^{t-1})\}, \epsilon_e^t, \hat{\Delta}T_e^t) \\ &= K_{common-sense}^{t-1} \cup \{(\epsilon_e^t, \omega_{acc}^t)\} \\ &\qquad \cup \{(\tau_x^{t-1}, \omega_x^{t-1} - \omega_{acc}^t)\} \end{aligned} \tag{11}$$

Where:
$$\omega_{acc} = \begin{cases} 1 - \hat{\Delta}T_e^t & \text{if first occurrence} \\ \theta_{memory} & \text{if reinforcement} \end{cases}$$

2. Episodic knowledge Episodic knowledge is updated in all cases when a new event is encountered. Given that the current event is e_t and the last event is e_{t-1}, the temporal sequence is established as $e_{t-1} \rightarrow e_t$. The strength of belief depends on whether this sequence is already remembered:

$$\begin{aligned} K_{episodic}^t &= f_{memory}(K_{episodic}^{t-1}, e_{t-1}, e_t, \theta_{episodic}) \\ &= K_{episodic}^{t-1} \cup \{((e_{t-1} \rightarrow e_t), \omega_{episodic})\} \end{aligned} \tag{12}$$



((a)) t_1: Frame 1 ((b)) t_2: Frame 2 ((c)) t_3: Frame 3

((d)) t_4: Frame 4 ((e)) t_5: Frame 5

Figure 2: Phase 1

Where:
$$\omega_{episodic} = \begin{cases} \theta_{memory} & \text{if first occurrence} \\ \omega_{(e_{t-1} \rightarrow e_t)}^{t-1} + \theta_{memory} & \text{if reinforcement} \end{cases}$$

3. visSit The infant shifts attention away from the current stimulus toward other objects or people if this infant experiences boredom from the perceived event.

6 Illustration

To illustrate how our model can be applied to concrete cases, we sketch an analysis of a sequence from the Rollins corpus involving a 9-month-old infant and his caregiver (see Figures 2, 3, and 4). In this interaction, the caregiver repeatedly performs a humorous action – throwing a toy circle from her head – which violates normal object use, creating a natural experiment in humour development. This example demonstrates how our formalism captures the dynamic process of humour recognition, from initial confusion to emerging laughter and anticipatory appreciation, illustrating the interplay between knowledge and contextual factors. Table 2 traces the infant's emotional and cognitive progression throughout this humorous interaction [5].

Context: The infant (with memory reinforcement parameter θ_{memory}) is playing with a toy

[5] A reviewer wonders how we determine the various thresholds discussed here. As the aim of this paper is to propose a formal model based on a principled, mechanistic-approach for early humorous interaction, the illustration in the text serves to demonstrate the model's capacity to explain the emergence of specific emotional responses in infants towards each caregiver-produced event. The thresholds for discrepancy values in this illustrative case study are therefore inferred from the observed infant's emotional responses (i.e., if a specific emotion expressing response like confusion, smile, or laughter is observed, we infer that the degree of discrepancy for that event falls within the corresponding threshold range for that emotion as defined in our model).

Appraisal Derivation	*Affect Derivation*			*Affect Consequence*		
	$emoExpProduced_t$	$pleasure_t$	$power_t$	$visSit_t$	$K^t_{common\text{-}sense}$	$K^t_{episodic}$
$\hat{\Delta T}^t_e \geq \theta_{fearLine}$	Fear	$f_{updatePleasure}$	$f_{updatePower}$	caregiver		f_{memory}
$\theta_{punchlineMax} < \hat{\Delta T}^t_e < \theta_{fearLine}$	Confusion				f_{res}	
$\hat{\Delta T}^t_e \approx \theta_{punchlineMax}$	Smile				f_{acc}	
$\theta_{punchlineMin} \leq \hat{\Delta T}^t_e \leq \theta_{punchlineMax}$	Laughter					
$\hat{\Delta T}^t_e < \theta_{punchlineMin}$	Boredom			¬ caregiver		

Table 1: Affect derived and knowledge update based on the appraisal value

Time	*Appraisal Derivation*	*Affect Derivation*			*Affect consequence*		
		$emoExpProduced_t$	$pleasure_t$	$power_t$	$visSit_t$	$K^t_{common\text{-}sense}$	$K^t_{episodic}$
t_1-t_2					¬ caregiver		
t_3	$\theta_{punchlineMax} < \hat{\Delta T}^{t_3}_{e_1} < \theta_{fearLine}$	Confusion	$f_{updatePleasure}$	$f_{updatePower}$	caregiver	f_{res}	f_{memory}
t_4	$\theta_{punchlineMax} < \hat{\Delta T}^{t_4}_{e_2} < \theta_{fearLine}$						
t_6	$\hat{\Delta T}^{t_6}_{e_1} \approx \theta_{punchlineMax}$	Smile				f_{acc}	
t_7	$\theta_{punchlineMin} \leq \hat{\Delta T}^{t_7}_{e_2} \leq \theta_{punchlineMax}$	Laughter					
t_8	$\theta_{punchlineMin} \leq \hat{\Delta T}^{t_8}_{e_{toyCirclePlay}} \leq \theta_{punchlineMax}$						

Table 2: The infant's emotional and cognitive progression during the toy circle play

bird in a relaxed state with the caregiver on the side watching him, indicating that both the environment and social presence are comfortable ($environment = 1, people = 1$), and from his neutral face, $pleasure = 0$.

6.1 Phase 1

Time t_1 (Frame 1): The clip begins with the infant playing with the toy bird while the caregiver observes him. They are not in intersubjective engagement as their gazes are not shared.

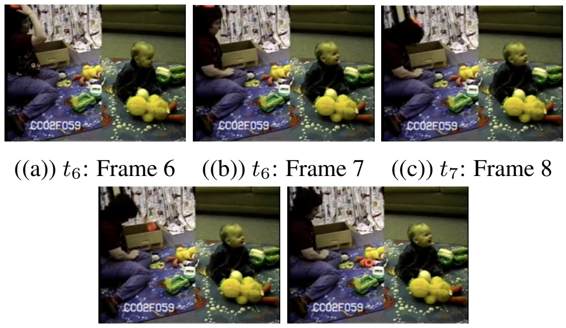
Time t_2 (Frame 2): The caregiver picks up a toy circle, though it does not yet attract the infant's attention.

Time t_3 (Frame 3): The caregiver places the toy circle on her head (e_1) with a neutral face, which attracts the infant's attention. At this point, they establish shared attention.

$$\text{caregiver.TIS} = \left[\text{public} : \begin{bmatrix} \text{time} & : t_3 \\ \text{visSit} & : baby \\ \text{eventProduced} & : \left[e_1 \begin{bmatrix} spkr & = \text{caregiver} \\ addr & = \text{baby} \\ e_{\text{action}} & = \text{places the toy Circle on the head} \\ e_{\text{prosodic}} & : \text{none} \\ e_{\text{face}} & : \text{neutral} \end{bmatrix} \right] \end{bmatrix} \right]$$

However, the infant's confused facial expression indicates that e_1 is not fully resolved and its discrepancy value within the confusion range ($\theta_{punchlineMax} < \hat{\Delta T}^{t_3}_{e_1} < \theta_{fearLine}$), with no contextual modulation from social cues. The infant attempts to bring about complete resolution to this event using existing knowledge and stores it in episodic memory.

Time t_4 (Frame 4): The caregiver then throws the toy circle from her head (e_2) with a neutral face, which continues to attract the infant's attention. The infant's continued confused expression indicates that e_2 is unresolved, with its discrepancy value in the confusion range ($\theta_{punchLineMax} <$



((a)) t_6: Frame 6 ((b)) t_6: Frame 7 ((c)) t_7: Frame 8

((d)) t_7: Frame 9 ((e)) Frame 10

Figure 3: Phase 2

$\hat{\Delta T}^{t_4}_{e_2} < \theta_{fearLine}$).

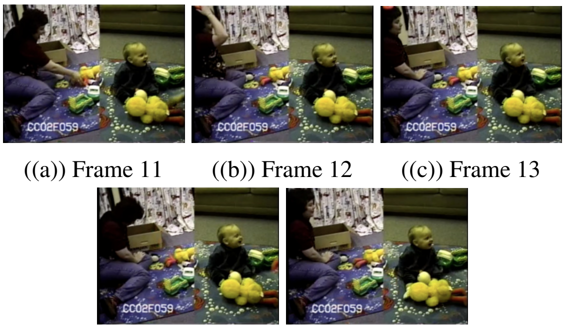
$$\text{caregiver.TIS} = \left[\text{public} : \begin{bmatrix} \text{time} & : t_4 \\ \text{visSit} & : \text{baby} \\ \text{eventProduced} & : \left[e_2 \begin{bmatrix} spkr & = \text{caregiver} \\ addr & = \text{baby} \\ e_{\text{action}} & = \text{throws toy Circle from her head} \\ e_{\text{prosodic}} & : \text{none} \\ e_{\text{face}} & : \text{neutral} \end{bmatrix} \right] \end{bmatrix} \right]$$

Time t_5 (Frame 5): The infant is looking at the caregiver's face, seeking social referencing cues. However, neither pitch contour nor facial expression from the caregiver provides sufficient information for the infant to resolve this sequence of events ($e_1 \rightarrow e_2$).

6.2 Phase 2

Time t_6 (Frame 6-7): The second phase starts. The caregiver places the toy circle on her head (e_1) again with neutral face. This time, the infant responds with a smile, indicating that the resolution process at t_3 has reduced the perceived discrepancy to approach the threshold for humour appreciation ($\hat{\Delta T}^{t_6}_{e_1} \approx \theta_{punchLineMax}$). The infant now begins to accommodate this new knowledge into his common-sense knowledge.

Time t_7 (Frame 8-9): The caregiver throws



((a)) Frame 11 ((b)) Frame 12 ((c)) Frame 13

((d)) Frame 14 ((e)) Frame 15

Figure 4: Phase 3

the toy circle from her head (e_2) and the infant responds with laughter – a clear sign that the resolution process at t_4 has successfully brought the discrepancy value into the optimal range for humour appreciation ($\theta_{punchLineMin} \leq \hat{\Delta}T_{e_2}^{t_7} \leq \theta_{punchLineMax}$). The infant now starts to accommodate this new learned knowledge into his common-sense knowledge.

6.3 Phase 3

Time t_8 (Frame 11-15): The caregiver merely picks up the toy circle, and remarkably, the infant begins laughing preemptively and continues throughout this phase. This anticipatory response demonstrates that the infant now perceives the entire sequence as a unified humorous routine ($e_{toyCirclePlay}$) rather than discrete actions. The temporal order of events ($e_1 \rightarrow e_2$) has been successfully encoded in episodic memory, enabling prediction and appreciation of the complete sequence.

7 Conclusion

This paper proposes a formal model of humour recognition in infant-caregiver interactions, specifically designed to address key limitations in current computational approaches to humour in dialogue systems, including limited explainability, inadequate consideration of context, and poor cognitive adaptability.

To address these challenges, we designed a formal model inspired by the OCC model's three-stage appraisal process and specified each appraisal stage based on existing psychological theories of early humour appreciation. This formal model offers enhanced explainability by explicitly tracing how an infant's specific emotional response arises from the interplay of three contextual factors (safety factors, social cues, and current emotional state) and their dynamic cognitive state (updates to common-sense and episodic knowledge during the interaction). This detailed mechanistic approach illustrates why the same event might elicit different emotional responses from an infant at different times, based on their evolving internal state.

We also introduce a novel approach to dialogue modeling by combining the KoS framework with TTR. We utilize KoS to construct a dual-state for each participant, representing both their *public* (external perceptions and productions) and *private* (internal knowledge and mood) states, exemplifying how two agents perceive, process, adapt, and respond to the communicative event their interlocutor produces. TTR, in turn, is employed to formally represent multimodal events, an agent's knowledge state, and their emotional state.

8 Limitations and future work

Despite these contributions, several limitations exist. While the case study effectively illustrates early humour appreciation, further validation with a broader range of examples is needed to refine the model's components and determine the necessary parameters (thresholds, contextual factor weights). Moreover, further work should explore the temporal relationship (Allen, 1983) between the caregiver's and the child's multimodal modalities (e.g., gaze patterns, facial expressions, and prosody) and their influence on the infant's emotional state.

Several promising directions for future research emerge from this work: 1) Expand the framework to include non-humorous laughables (events that trigger laughter without humour, such as uses of laughter to respond to another's friendliness (El Haddad et al., 2019; Mazzocconi et al., 2023) or establish shared attention with others (Mazzocconi and Ginzburg, 2022; Parnell, 2023)); 2) Extend our model to incorporate specific characteristics of caregiver-infant interactions, such as interactional synchrony patterns and emotion regulation mechanisms (Feldman, 2007); 3) Implement our formal model in a robotic platform where the robot would act as an infant responding to humorous stimuli, enabling experimental validation of humour appreciation theories, controlled parameter testing, and refinement of the theoretical framework.

Acknowledgments

This work was supported by the French National Centre for Scientific Research (CNRS) under the reference 80PRIME2023—TELIN and by the investment program "France 2030" launched by the French Government and implemented by the University Paris Cité as part of its program "Initiative d'excellence" IdEx with the reference (ANR-10-IDEX-0001). We would like to thank three SiGDial 2025 reviewers for their valuable comments on the first draft of this paper.

References

James F Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Renee Baillargeon, Elizabeth S Spelke, and Stanley Wasserman. 1985. Object permanence in five-month-old infants. *Cognition*, 20(3):191–208.

Daniel E Berlyne. 1960. Conflict, arousal, and curiosity.

Gustavo Betarte, Alvaro Tasistro, G Sambin, and J Smith. 1998. Martin-löf's type theory with record types and subtyping. *G. Sambin & J. Smith (eds.)*, 25.

Hanne C Braarud and Kjell Morten Stormark. 2006. Expression of negative affect during face-to-face interaction: a double video study of young infants' sensitivity to social contingency. *Infant and Child Development: An International Journal of Research and Practice*, 15(3):251–262.

Ellen Breitholtz. 2014. Enthymemes in dialogue: A micro-rhetorical approach.

Ellen Breitholtz. 2020. *Enthymemes and Topoi in Dialogue: the use of common sense reasoning in conversation*. Brill.

Ellen Breitholtz and Robin Cooper. 2011. Enthymemes as rhetorical resources. *Proceedings of, volume SemDial*, pages 149–157.

Steven L Bressler and Vinod Menon. 2010. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, 14(6):277–290.

Gordon W Bronson. 1972. Infants' reactions to unfamiliar persons and novel objects. *Monographs of the Society for Research in Child Development*, pages 1–46.

Richard L. Canfield and Marshall M. Haith. 1991. Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, 27:198–208.

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3:333–362.

Robin Cooper. 2012. Type theory and semantics in flux. *Handbook of the Philosophy of Science*, 14:271–323.

Robin Cooper. 2023. *From perception to communication: a theory of types for action and meaning*. Oxford University Press.

Robin Cooper and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. *The handbook of contemporary semantic theory*, pages 375–407.

Kevin El Haddad, Sandeep Nallan Chakravarthula, and James Kennedy. 2019. Smile and laugh dynamics in naturalistic dyadic interactions: Intensity levels, sequences and roles. In *2019 international conference on multimodal interaction*, pages 259–263.

Magy Seif El-Nasr, John Yen, and Thomas R Ioerger. 2000. Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3:219–257.

Ruth Feldman. 2007. Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child psychology and Psychiatry*, 48(3-4):329–354.

Raquel Fernández Rovira. 2006. *Non-sentential utterances in dialogue: Classification, resolution and use*. Ph.D. thesis, King's College London (University of London).

Sigmund Freud. 1960. *Jokes and their relation to the unconscious*. Routledge, & KaganPaul.

Patrick Gebhard. 2005. Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36.

Jonathan Ginzburg. 1994. An update semantics for dialogue. In H. Bunt, editor, *Proceedings of the 1st International Workshop on Computational Semantics*. ITK, Tilburg University, Tilburg.

Jonathan Ginzburg. 2012. *The interactive stance*. Oxford University Press, USA.

Jonathan Ginzburg. 2016. The semantics of dialogue. *The Cambridge handbook of formal semantics*.

Jonathan Ginzburg and Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and philosophy*, 27:297–365.

Jonathan Ginzburg and Raquel Fernández. 2010. Computational models of dialogue. *The handbook of computational linguistics and natural language processing*, pages 429–481.

Jonathan Ginzburg, Chiara Mazzocconi, and Ye Tian. 2020. Laughter as language. *Glossa: a journal of general linguistics (2021-...)*, 5(1).

Elena Hoicka and Merideth Gattis. 2008. Do the wrong thing: How toddlers tell a joke from a mistake. *Cognitive Development*, 23(1):180–190.

Elena Hoicka and Merideth Gattis. 2012. Acoustic differences between humorous and sincere communicative intentions. *British Journal of Developmental Psychology*, 30(4):531–549.

Elena Hoicka, Burcu Soy Telli, Eloise Prouten, George Leckie, William J Browne, Gina Mireault, and Claire Fox. 2022. The early humor survey (ehs): A reliable parent-report measure of humor development for 1-to 47-month-olds. *Behavior research methods*, pages 1–26.

Dianne Horgan. 1981. Learning to tell jokes: A case study of metalinguistic abilities. *Journal of Child Language*, 8(1):217–224.

Yingqin Hu, Capucine Brillet, Bosko Rajkovic, Gauhar Rustamova, Chiara Mazzocconi, Catherine Pelachaud, and Jonathan Ginzburg. 2024. Laughter in the cradle: A taxonomy of infant laughables. In *Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue*.

Moritz Köster, Ezgi Kayhan, Miriam Langeloh, and Stefanie Hoehl. 2020. Making sense of the world: Infant learning from a predictive processing perspective. *Perspectives on psychological science*, 15(3):562–571.

Staffan Larsson. 2002. *Issue-based dialogue management*. Citeseer.

Maria Legerstee. 1997. Contingency effects of people and objects on subsequent cognitive functioning in three-month-old infants. *Social development*, 6(3):307–321.

Andy Lücking and Jonathan Ginzburg. 2023. Leading voices: Dialogue semantics, cognitive science and the polyphonic structure of multimodal interaction. *Language and Cognition*, 15(1):148–172.

Stacy C Marsella and Jonathan Gratch. 2009. Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90.

Chiara Mazzocconi, Kevin El Haddad, Benjamin O'Brien, Kübra Bodur, and Abdellah Fourtassi. 2023. Laughter mimicry in parent-child and parent-adult interaction. In *Multimodal Communication Symposium 2023*.

Chiara Mazzocconi and Jonathan Ginzburg. 2022. A longitudinal characterization of typical laughter development in mother–child interaction from 12 to 36 months: Formal features and reciprocal responsiveness. *Journal of Nonverbal Behavior*, 46(4):327–362.

Chiara Mazzocconi, Ye Tian, and Jonathan Ginzburg. 2020. What's your laughter doing there? a taxonomy of the pragmatic functions of laughter. *IEEE Transactions on Affective Computing*, 13(3):1302–1321.

Robert B McCall and Paul E McGhee. 1977. The discrepancy hypothesis of attention and affect in infants. In *The structuring of experience*, pages 179–210. Springer.

Paul E McGhee. 1976. Children's appreciation of humor: A test of the cognitive congruency principle. *Child Development*, pages 420–426.

Paul E McGhee and Edie Pistolesi. 1979. *Humor: Its origin and development*. WH Freeman San Francisco.

Gina Mireault, John Sparrow, Merlin Poutre, Brittany Perdue, and Laura Macke. 2012. Infant humor perception from 3-to 6-months and attachment at one year. *Infant Behavior and Development*, 35(4):797–802.

Gina C Mireault, Susan C Crockenberg, John E Sparrow, Kassandra Cousineau, Christine Pettinato, and Kelly Woodard. 2015. Laughing matters: Infant humor in the context of parental affect. *Journal of Experimental Child Psychology*, 136:30–41.

Gina C Mireault, Susan C Crockenberg, John E Sparrow, Christine A Pettinato, Kelly C Woodard, and Kirsten Malzac. 2014. Social looking, social referencing and humor perception in 6-and-12-month-old infants. *Infant Behavior and Development*, 37(4):536–545.

Gina C Mireault and Vasudevi Reddy. 2020. Making sense of infants' differential responses to incongruity. *Human development*, 64(2):55–63.

Donna L Mumme, Anne Fernald, and Carla Herrera. 1996. Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child development*, 67(6):3219–3237.

Andrew Ortony, Gerald L Clore, and Allan Collins. 2022. *The cognitive structure of emotions*. Cambridge university press.

Vicki Parnell. 2023. " watch me, mom!" the development of infants' skills in eliciting others' attention.

Jean Piaget, Margaret Cook, and 1 others. 1952. *The origins of intelligence in children*, volume 8. International universities press New York.

Matthew Purver. 2006. Clarie: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4:259–288.

Aarne Ranta. 1995. *Type-theoretical grammar*. Oxford University Press.

Victor Raskin. 1979. Semantic mechanisms of humor. In *Annual Meeting of the Berkeley Linguistics Society*, pages 325–335.

Vasudevi Reddy. 1991. Playing with others' expectations: Teasing and mucking about in the first year.

Vasudevi Reddy and Gina Mireault. 2015. Teasing and clowning in infancy. *Current Biology*, 25(1):R20–R23.

Vasudevi Reddy 1. 2001. Infant clowns: The interpersonal creation of humour in infancy. *Enfance*, 533(3):247–256.

Farfalla Ribordy, Adeline Jabès, Pamela Banta Lavenex, and Pierre Lavenex. 2013. Development of allocentric spatial memory abilities in children from 18 months to 5 years of age. *Cognitive psychology*, 66(1):1–29.

Hannes Ritschel, Thomas Kiderle, Klaus Weber, Florian Lingenfelser, Tobias Baur, and Elisabeth André. 2020. Multimodal joke generation and paralinguistic personalization for a socially-aware robot. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness. The PAAMS Collection: 18th International Conference, PAAMS 2020, L'Aquila, Italy, October 7–9, 2020, Proceedings 18*, pages 278–290. Springer.

PR Rollins. 2003. Caregiver contingent comments and subsequent vocabulary comprehension. *Applied Psycholinguistics*, 24(2):221–234.

Mary K Rothbart. 1973. Laughter in young children. *Psychological bulletin*, 80(3):247.

David C Rubin. 2000. The distribution of early childhood memories. *Memory*, 8(4):265–269.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Jenny R Saffran, Richard N Aslin, and Elissa L Newport. 1996. Statistical learning by 8-month-old infants. *science*, 274(5294):1926–1928.

Thomas R Shultz. 1972. The role of incongruity and resolution in children's appreciation of cartoon humor. *Journal of Experimental Child Psychology*, 13(3):456–477.

Thomas R Shultz. 1974. Development of the appreciation of riddles. *Child development*, pages 100–105.

Thomas R Shultz. 2017. A cognitive-developmental analysis of humour. In *Humor and laughter*, pages 11–36. Routledge.

James F Sorce, Robert N Emde, Joseph J Campos, and Mary D Klinnert. 1985. Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. *Developmental psychology*, 21(1):195.

Robert Soussignan, Jacqueline Nadel, Pamela Canet, and Priscille Gerardin. 2006. Sensitivity to social contingency and positive emotion in 2-month-olds. *Infancy*, 10(2):123–144.

Herbert Spencer. 1875. The physiology of laughter.

L Alan Sroufe and Jane Piccard Wunsch. 1972. The development of laughter in the first year of life. *Child development*, pages 1326–1344.

LA Sroufe, E Waters, and L Matas. 1974. Contextual determinants of infant affective response. *The origins of fear*, 2:49–72.

Margaret Wolan Sullivan and Michael Lewis. 1989. Emotion and cognition in infancy: Facial expressions during contingency learning. *International Journal of Behavioral Development*, 12(2):221–237.

Jerry M Suls. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues*, 1:81–100.

Xiao Sun, Xiaoqi Peng, and Shuai Ding. 2018. Emotional human-machine conversation generation based on long short-term memory. *Cognitive Computation*, 10(3):389–397.

Lev Semenovich Vygotsky and Michael Cole. 1978. *Mind in society: Development of higher psychological processes*. Harvard university press.

Klaus Weber, Hannes Ritschel, Ilhan Aslan, Florian Lingenfelser, and Elisabeth André. 2018. How to shape the humor of a robot-social behavior adaptation based on reinforcement learning. In *Proceedings of the 20th ACM international conference on multimodal interaction*, pages 154–162.

Kaibin Xu, Junpei Zhong, and Kristiina Jokinen. 2021. It is time to laugh: Discovering specific contexts for laughter with attention mechanism. In *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pages 211–215. IEEE.

A Appendix

A.1 TTR

Type Theory with Records (TTR) (Cooper, 2005, 2012; Cooper and Ginzburg, 2015; Cooper, 2023) constitutes a formal semantic framework that synthesizes the proof-theoretic mechanisms of Martin-Löf Type Theory (Ranta, 1995; Betarte et al., 1998) with the situation-theoretic approach to meaning representation.

Records: The record A.1 in TTR is an attribute-value matrix (AVM) that assigns specific values to a set of labels, which is used to define an event or state.

$$\begin{bmatrix} label_1 & = value_1 \\ label_2 & = value_2 \\ ... & = ... \\ label_n & = value_n \end{bmatrix}$$

Record types : Record types are sets of label/type pairs, each representing a judgement, the types can embed other types, in particular record types.

$$\begin{bmatrix} label_1 & : Type_1 \\ label_2 & : Type_2 \\ ... & : ... \\ label_n & : Type_n \end{bmatrix}$$

The record $\begin{bmatrix} label_1 & = value_1 \\ label_2 & = value_2 \\ ... & = ... \\ label_n & = value_n \end{bmatrix}$ is of type $\begin{bmatrix} label_1 & : Type_1 \\ label_2 & : Type_2 \\ ... & : ... \\ label_n & : Type_n \end{bmatrix}$ iff $value_1 : Type_1, value_2 : Type_2, ..., value_n : Type_n$

The type judgement: For the same object o, the agent A classifies the object o as the Type T_1 $o :_A T_1$, while the agent A classifies the object o as the Type T_2 $o :_B T_2$.

A.2 Topoi and enthymemes

The representation of topoi and enthymemes in TTR is extended by Breitholtz and Cooper (2011); Breitholtz (2014, 2020).

Topoi A topos (τ) is defined as a reasoning pattern and formalized as a function from one situation type to another situation type ($\tau : RecType \to RecType$), and represents one's private reasoning principles that guide expectations ($\tau = \lambda e : D_1 \cdot R_1$). D_1 represents the domain conditions and R_1 represents the expected range or outcome.

$$\tau = \lambda e.[D_1 : RecType] \cdot [R_1 : RecType] \quad (13)$$

Enthymemes An enthymeme (ϵ) is defined as the argument derived from the actual conversation and also formalized as a function from one situation type to another situation type ($\epsilon : RecType \to RecType$). Enthymemes represent specific instances of reasoning that may or may not align with established topoi ($\epsilon = \lambda e : D_2 \cdot R_2$).

$$\epsilon = \lambda e.[D_2 : RecType] \cdot [R_2 : RecType] \quad (14)$$

Therefore, given the fact that the Line-1 is faster than Line-2, for an agent who needs to arrive at a destination in a timely manner, choosing Line-1 represents a situation where the enthymeme conforms to the topos T_{faster} (given two metro lines, choose the faster one).

$$\lambda r. \begin{bmatrix} x & : \text{Ind} \\ y & : \text{Ind} \\ c_{\text{metro1}} & : \text{metro}(\text{x}) \\ c_{\text{metro2}} & : \text{metro}(\text{y}) \\ c_{\text{faster}} & : \text{faster}(\text{x,y}) \end{bmatrix} . \begin{bmatrix} c_{\text{choose}} & : \text{choose}(\text{r.x}) \end{bmatrix}$$

$$\lambda r. \begin{bmatrix} x & = \text{Line-1} \\ y & = \text{Line-2} \\ c_{\text{metro1}} & : \text{metro}(\text{x}) \\ c_{\text{metro2}} & : \text{metro}(\text{y}) \\ c_{\text{faster}} & : \text{faster}(\text{x,y}) \end{bmatrix} . \begin{bmatrix} c_{\text{choose}} & : \text{choose}(\text{r.y}) \end{bmatrix}$$

A.3 KoS

Our formalisation of interaction is formulated within the framework KoS (Ginzburg, 1994; Larsson, 2002; Purver, 2006; Fernández Rovira, 2006; Ginzburg and Fernández, 2010; Ginzburg, 2012; Ginzburg et al., 2020).

KoS provides an approach to model an agent's dynamic cognitive states (a private and a dialogue gameboard (public)), including their beliefs, questions currently at issue, their visual focus of attention, salient utterances and other relevant information in a conversation.

$$TotalCognitiveState =_{def} \begin{bmatrix} \text{DIALOGUEGAMEBOARD} & : \mathit{DGBType} \\ \text{PRIVATE} & : \mathit{private} \end{bmatrix}$$

The dialogue gameboard (DGB) records information from each agent's perspective. It keeps track of turn holders (speaker (spkr), addressee (addr)), the shared assumptions (FACT), the questions under discussion (QUD), the interaction history, both grounded and ungrounded, respectively (Moves and Pending), the visual situation (VisSit), and the participant's public emotional state, Mood.

$$DGBType =_{def}$$

$$\begin{bmatrix} \text{spkr} & : \text{Ind} \\ \text{addr} & : \text{Ind} \\ \text{utt-time} & : \text{Time} \\ \text{c-utt} & : \text{addressing}\big(\text{spkr, addr, utte-time}\big) \\ \text{Facts} & : \text{set}\big(\text{propositions}\big) \\ \text{VisSit} & : \big[\text{InAttention} \ : \text{Ind}\big] \\ \text{Pending} & : \text{list}\big(\text{LocProp}\big) \\ \text{Moves} & : \text{list}\big(\text{IllocProp}\big) \\ \text{QUD} & : \text{poset}\big(\text{Question}\big) \\ \text{Mood} & : \text{Appraisal} \end{bmatrix}$$

A.3.1 Conversational rules

Conversational rules define the mapping actions between two cognitive states (preconditions and effects), meaning if the current case satisfies the range of cases (preconditions), then one can update as specified in the effects.