Investigating Narratives of Social Intention in Restaurant Interactions

Researching scenarios for Intention Prediction

Aleksander Sak[1]

Supervisor(s): Hayley Hung[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 25, 2026

Name of the student: Aleksander Sak
Final project course: CSE3000 Research Project
Thesis committee: Hayley Hung, Hayley Hung, Ricardo Marroquim

Abstract

Inferring social intention in everyday settings is challenging because the same observable behavior can support multiple plausible interpretations. This issue is pronounced in restaurants, where roles and norms structure interaction but do not uniquely determine what is socially "meant." The research question addressed is: How can scenarios in a restaurant setting be created that allow investigation into how humans and intelligent systems construct multiple plausible narratives of social intention?

A literature-grounded scenario design method is presented that separates observable cues from inferred psychological meanings and situation classes, and frames interactions using external scripts (norms, roles, scenes) and internal script variants (observer-dependent interpretations). Two short scenarios with three controlled variations each are specified to modulate openendedness through cue completeness, norm clarity, and perspective. It is concluded that openendedness can be designed systematically by controlling observability and introducing norm tensions and perspective differences that keep multiple narratives simultaneously plausible.

1 Introduction

Understanding and inferring social intention is a central challenge for intelligent systems operating in human social environments. In domains such as restaurants, where interactions are shaped by social norms, roles, and expectations, observable behavior often under-specifies intent. An identical action, such as waiting, approaching a table, or making eye contact, may give rise to multiple plausible interpretations depending on the observer's perspective, prior experiences, and assumptions about the situation. Designing research that meaningfully evaluates an intelligent system's ability to infer social intention in such settings therefore raises a question: How to create scenarios in restaurant setting that allow investigation into how humans and intelligent systems construct multiple plausible narratives of social intention?

This question becomes more urgent in light of recent critique of mainstream intention estimation practice. Hung et al.[8] argue that many state-of-the-art approaches implicitly treat intention as equivalent to the realized outcome, which they describe as the *intention-by-outcome* problem. In this framing, intentions risk being treated as "valid" only when they come true, while unrealized intentions and competing goals in multi-person settings are sidelined. They further highlight that labeling intentions in the absence of outcomes is difficult and that third-party observer labels may produce multiple plausible narratives shaped by observers' experiences rather than a single ground truth. This motivates a shift from treating disagreement as noise to treating it as informative variation that research should explicitly represent.

A complementary motivation comes from work on interpersonal causal attribution and how it changes when new information becomes available. Michener et al.[13] study retrospective revaluation in social attribution, showing that when two candidate causes are initially plausible, later information that weakens one cause can systematically increase attribution to the other (with deflation effects reported as stronger than inflation effects). This demonstrates that social explanations are not only subjective, but also dynamically updated as additional cues or contextual information are introduced. For scenario-based intention research, this provides a concrete rationale for designing interactions where ambiguity is not static: later cues can rationally shift which narrative becomes most plausible.

The goal of this paper is to explore how scenarios can be constructed in a restaurant setting to systematically investigate variation in perceived social intention. The supporting questions are: (1) Which dimensions of restaurant situations and observer perspectives are most likely to produce diverging yet plausible intention narratives? (2) How can scenarios be structured so that an intelligent system is evaluated on producing and explaining multiple plausible narratives?

This paper contributes: (i) a structured literature survey that extracts narrative targets, divergence triggers, and privacy-constrained cue sets; (ii) a representation bridge for scenario design that separates observable cues from inferred psychological characteristics and situation classes; and (iii) two compact restaurant interaction case studies with controlled variations to modulate openendedness in intention interpretation.

The remainder of the paper is structured as follows. Section 2 describes the methodology for conducting the literature study and translating its outputs into scenario case studies. Section 3 presents the literature survey and derives scenario design variables grounded in prior work. Section 4 introduces the scenario case studies and their variations using a consistent template. Later sections discuss responsible research considerations, interpret results in relation to the literature, and conclude with limitations and recommendations for future work.

2 Methodology

This section describes how the literature study was conducted and how its outputs were operationalized into compact, openended restaurant interaction scenarios. The methodology proceeds in two stages. First, a literature study was performed to identify (i) what cues can realistically be sensed in restaurant-like settings, (ii) which psychological meanings and narrative constructs are commonly inferred from those cues, and (iii) how restaurant interactions can be represented as externally scripted norms versus internally configured interpretations. Second, these findings were translated into two short case studies (scenarios), each instantiated in three controlled variations to systematically modulate openendedness in intention interpretation.

2.1 Literature study

After establishing the motivation of investigating *perceived* social intention and defining the goal of the literature study, a structured query was constructed to retrieve candidate papers via Scopus (the full query is provided in appendix A).

The Scopus search returned 123 papers in total. Titles and abstracts were screened for relevance using criteria aligned with the project scope, namely: the presence of restaurant or restaurant-like social settings, an explicit focus on perception, intention, or social interpretation, and the contribution of either computational methods, datasets, or theoretical structure useful for modeling interpretive variation. This process yielded 16 relevant papers.

To expand coverage beyond the initial search results, snowballing was applied to the selected papers using backward and forward reference tracing. This produced an additional 3 relevant works, resulting in a final set of 19 papers used in the literature survey.

The 19 papers were utilized in Section 3 by clustering them into three groups: (1) social perception and behavioral intention in restaurant interactions, (2) perception and activity understanding in restaurant-like scenes under privacy constraints, and (3) intent modeling for conversational and multi-domain systems. Each cluster was interpreted through a representation bridge that distinguishes cues (observable elements), characteristics (inferred psychological meanings that can vary across observers), and classes (abstract situation types), and was linked to script theory terminology (play, scene, role, scriptlet; internal vs. external scripts). The main outputs of the literature survey used for scenario creation were: narrative targets (e.g., warmth/competence, trustworthiness, sincerity/relief, reciprocity, decision difficulty, satisfaction leading to behavioral intention), divergence triggers (e.g., norm violations, attribution of responsibility, preference communication moves, incentive and payment rules), and a compact list of scenario variables organized as cue-layer variables, external-script variables, and internal-script variables, together with a reusable scenario template.

2.2 Case study development (scenario creation)

Scenario creation was conducted as the development of short, interaction-focused case studies rather than full restaurant visits. Each case study was specified using a structured template that separates: (i) cues that are physically present and potentially observable under privacy constraints, (ii) the external script that describes the normative play and expected scriptlets in the scene, and (iii) internal script variants that represent alternative plausible observer interpretations driven by differences in goals and perceived situational characteristics. This operationalization follows the idea that the same cues can support multiple plausible narratives, and that openendedness increases when norms are weak or disrupted, cues are incomplete, or internal scripts differ.

Two scenarios were created, each with three variations. Variations are defined as different versions of the same interaction rather than new scenario classes. Following the design guidance used in this project, each scenario includes at least one person for whom intention interpretation exhibits at least two levels of openendedness, achieved by varying at least two dimensions (for example: responsibility clarity and response style; or norm clarity and cue completeness). One scenario was designed to expand an existing line of work identified in the literature survey, while the other was designed to address a gap by combining constructs that were previously studied separately into a single compact interaction that still preserves privacy-constrained observability.

For each scenario and each variation, the construction procedure was:

1. **Select a class (situation type)** grounded in the literature clusters (e.g., service recovery; payment coordination).
2. **Define the cue layer** as a minimal set of observable elements (persons/roles present, objects such as bill or cup, and short activity traces such as reaching, pausing, replacing), while explicitly stating the allowed observation policy (e.g., video-only activity and object-state cues, no audio).
3. **Specify the external script** by writing the normative play and the local scene structure (roles, expectations, and scriptlets), then introduce a controlled disruption (failure, norm tension, under-specification, or incentive tension).
4. **Enumerate internal script variants** (at least two) by varying observer goals and perceived situational characteristics, yielding multiple plausible intention narratives for the same observed cues.
5. **Identify divergence points** where cues are ambiguous or under specified (occlusion, unclear responsibility, mixed norms, role uncertainty, or strategic communication), and where different internal scripts would plausibly be activated.
6. **Define personalization and tuning variables** that modify openendedness without changing the scenario class (e.g., relationship type, cultural expectations about norms, prior familiarity with robots, visibility of consumption asymmetry), and specify which cues can be revealed or masked to control interpretive breadth.

2.3 Analysis, discussion, and future recommendations

The Results section reports the two scenarios and their variations using the scenario template developed in Section 3. The Discussion then interprets how each variation differentiates in cue availability, external-script strength, and internal-script configuration, and relates these choices back to the narrative targets and divergence triggers identified in the literature survey. Finally, future recommendations are formulated as design guidance for constructing scenario-based evaluations that preserve multiple plausible narratives of perceived social intention rather than enforcing a single consensus label.

3 Literature Survey

To ensure that this literature survey directly supports scenario construction (Section 4), each reviewed work is interpreted as contributing building blocks at three levels of situational description: (i) *cues* (physically present, potentially machine-detectable elements such as persons, activities, events, objects, location, and time), (ii) *characteristics* (psychological meanings inferred from cues, which may vary across observers), and (iii) *classes* (abstract situation types). This cues/characteristics/classes distinction is useful because cues alone do not determine what is "psychologically going on". The same cue constellation can support different meanings for different perceivers [17].

We connect these situation descriptors to script theory, treating restaurant interactions as plays composed of scenes, roles, and scriptlets. In this view, *external scripts* describe the normative version of the situation, while *internal scripts* represent an individual's working understanding of "what is happening." Internal scripts are dynamically configured and reconfigured, and their configuration is influenced by an observer's goals and their perceived situational characteristics [6]. This scaffolding motivates our emphasis on multiple plausible narratives of perceived social intention rather than a single consensus label.

3.1 Social perception and behavioral intention in restaurant interactions

Restaurant workplace studies define perceived social intention through the warmth and competence dimensions, and relate these evaluations to employees' attitudes and retention outcomes. Bufquin et al.[4] report that restaurant employees' evaluations of coworkers and managers clustered into two dominant profiles, warm-and-competent versus cold-and-incompetent, rather than a richer typology. These profiles were associated with systematic differences in job satisfaction, organizational commitment, and turnover intentions (for both coworker and manager targets).

Complementing this, Bufquin et al.[2] examine social evaluations of restaurant managers and find that perceived warmth and competence did not emerge as distinct constructs in their measurement model, leading to a single factor labeled "manager's warmth and competence." The paper then relates this construct to job satisfaction, organizational commitment, and turnover intentions via a multilevel path model.

A related study links these social perceptions to operational outcomes. Bufquin et al.[3] use structural equation modeling to examine effects of employees' social perceptions (coworkers and managers) and affective organizational commitment on restaurant performance indicators, including customer satisfaction, sales change, and turnover rate. Reported results emphasize that managers' warmth and employees' organizational commitment have significant effects on turnover rates, and that organizational commitment improves customer satisfaction and sales change.

Several papers show that small, interpretable deviations from norms can shift observer narratives about intention-relevant traits. Alley's restaurant video study[1] manipulates a brief incident of minor food theft and finds large changes in perceived trustworthiness and "sneakiness" for perpetrators, alongside changes in perceived self-confidence (generally higher for thieves and lower for victims), and reports downstream consequences such as altered hypothetical hiring decisions and inferred likelihood of romantic infidelity.

Normative expectations also shape how service recovery behavior is interpreted, including in human-robot service settings. Shan et al.[19] manipulate symbolic service recovery style (humorous vs. rational response) and robot social perception (warm vs. competent) and report an interaction on service evaluations: warm-perceived robots are evaluated more positively with humorous recovery, whereas competent-perceived robots are evaluated more positively with rational recovery. The paper further tests perceived relief and perceived sincerity as mediators, and locus of responsibility (robot error vs. customer error) as a boundary condition.

Customer reciprocity and behavioral intention are also modeled as responses to perceived relationship investment and experience quality. Ryu and Lee [18] split upscale restaurant customers into high versus low perceived relationship marketing investment groups and report broad differences across perceived restaurant quality dimensions (price fairness, food quality, service quality, physical environment), relational benefits (confidence, social, special treatment), and reciprocity proxies (revisit intentions and favorable reciprocal behaviors). They report that service quality, confidence benefits, and favorable reciprocal behaviors most strongly discriminate high- versus low-RMI customers.

Şahinoğlu and Başar[23] model restaurant experience, satisfaction, and behavioral intention, and report a regression in which customer satisfaction strongly predicts behavioral intention (with reported $R^2 \approx 0.927$, interpreted as explaining 92% of variance in behavioral intention).

Kadavil and Usha[9] study sensory brand recognition in quick-service restaurants and frame sensory inputs (visual, olfactory, gustatory, tactile, auditory) as drivers of brand awareness and downstream loyalty. In their reported conclusions, visual appeal and smell are emphasized for creating brand awareness, and furniture quality, spatial layout, and food flavor are reported as influential for brand loyalty, whereas background music/sounds are reported as not significant in their findings.

Two papers treat restaurant decisions as joint-choice settings where communication and incentives alter perceived intentions and outcomes. Kim et al.[12] examine "no-preference communication" in joint decisions (including choosing a restaurant) and report that recipients infer undisclosed preferences, experience greater decision difficulty, like the co-consumer less, shift choices away from their own most-preferred option, and enjoy the joint consumption less. They also report that communicators do not anticipate these negative effects. Gneezy et al.[7] study bill payment rules and report that diners consume more under even-split payment than individual pay, producing inefficiency, while many diners report preferring the individual-pay rule.

Scenario construction takeaways (Cluster 1). Across these studies, the *narrative targets* (characteristics) used to describe perceived intention and its consequences include warmth/competence impressions (or a combined warmth-competence factor), trustworthiness/sneakiness, perceived sincerity/relief in recovery, reciprocity intentions, satisfaction leading to behavioral intention, co-consumer liking, and perceived decision difficulty.

Key *divergence triggers* and normative leverage points studied include: norm violations (taking food without permission), attribution of responsibility for failure in service recovery, communication moves that conceal or under-specify preferences ("no preference"), and incentive structure shifts (even-split vs. individual pay).

These map onto common restaurant *plays/scenes* discussed or implied by the papers, including: dining interaction during eating, service failure and recovery, joint decision making about venue or options, and paying.

3.2 Perception and activity understanding in restaurant-like scenes

A second cluster constrains what can be sensed in restaurant settings and proposes datasets and models for recognizing activities and group states under realistic limitations.

Kaiser et al.[10] introduce the Noldus Database for automated recognition of restaurant-related activities in a "Restaurant of the Future" context, motivated by heavy reliance on manual video annotation and an explicit dismissal of acoustic cues for privacy reasons. The dataset comprises 11 videos of approximately 4 minutes each (about 68,000 frames total) with frame-level labels for activities such as coming in, sitting, drinking, eating, and going out. The paper also reports baseline recognition modules and example robustness metrics for tracked objects (for example, plate detection around 91.8% and glass tracking around 83.0%).

Taylor et al.[22] study group activity recognition for restaurant service from single-viewpoint footage using publicly available restaurant webcam videos of five two-person meals. They emphasize that meal phases and "neediness" require modeling an underlying table-level group state, not only individual actions, and describe preliminary classification efforts over individual activities and group states.

Kim et al.[11] propose Café, a multi-person group activity detection benchmark recorded in six cafes with four camera viewpoints. They report more than four hours of video segmented into 6-second clips and provide rich annotations including 3.5M human bounding boxes, track IDs, group configurations, and group activity labels. They also report that clips include multiple groups and many outliers, and that the number of actors per clip ranges from 3 to 14. Alongside the dataset, the paper proposes a Transformer-based model designed to localize an unknown number of groups and classify group activities.

Quiroz et al.[16] address group emotion detection from a robocentric perspective. Their pipeline detects faces, recognizes individual emotions (using VGGFace-based processing), aggregates to frame emotion and then to scene-level group emotion, and proposes dataset creation in simulated ROS/Gazebo environments (cafeteria and museum). Reported results include high individual emotion detection accuracy and group (scene) emotion detection per-frame accuracy around 90.84% in the cafeteria and 89.78% in the museum scenarios.

Mokhtari et al.[14] propose experience-based planning domains (EBPDs) where robots acquire episodic experiences, conceptualize them into activity schemata via generalization and feature extraction, and exploit schemata for planning. The paper integrates explicit goal inference, describing a procedure that extracts goal propositions from key propositions and task arguments, and reports evaluation in restaurant scenarios in both Gazebo simulation and on a PR2 robot.

Scenario construction takeaways (Cluster 2). These works collectively specify feasible *cue sets* for privacy-constrained restaurant sensing, including activity primitives (enter, sit, drink, eat, leave), object states (plate, glass), group configuration and group state descriptors (table-level meal phase and "neediness"), multi-group structure with outliers, and optionally aggregate affect at the group or scene level.

They also document natural *ambiguity sources* that can support multiple narratives: partial observability and occlusion (hands covering objects, viewpoint changes), dense scenes with many outliers, and role or group assignment uncertainty (who belongs to which group, who initiates an action), as well as the constraint of single-view versus multi-view sensing.

3.3 Intent modeling for conversational and multi-domain intelligent systems

A third cluster focuses on inferring structured intent representations from language and interaction traces, including restaurant booking, culinary chatbots, data augmentation for intent and slot learning, and weak supervision for complex intent discovery.

Fernando and Ganegoda[5] present ResBot, a bilingual (Sinhala and English) restaurant booking system built in Rasa. They propose a hybrid intent classification strategy in which DIET is used for intent and entity recognition, and a zero-shot intent classifier (using a BART model) is invoked when DIET confidence falls below a threshold; if confidence remains insufficient, the system asks the user to rephrase. The paper also describes a language handler using language detection and translation to English for consistent NLU processing, and reports evaluation comparing default versus best pipeline configurations.

Sofia et al.[20] incorporate IndoBERT into the DIET classifier for an Indonesian culinary chatbot, motivated by non-standard and slang language. They report a dataset of 4,011 sentences across 17 intents plus scraped restaurant/menu entities (1,079 menus from 203 restaurants), and compare ML (SVM+CRF), DIET, and DIET+IndoBERT pipelines. Reported results show higher F1 for DIET+IndoBERT (around 0.92 for intent and 0.96 for entity recognition) than DIET and the ML baseline.

Papangelis et al.[15] propose Generative Conversational Networks, a meta-learning framework where a generator produces labeled training data from seed data and a learner is trained on generated data, using reinforcement learning (PPO) to optimize generation by downstream task reward. They evaluate on intent detection and slot tagging across benchmarks including Restaurants8k, and report large average improvements over a baseline trained only on seed data.

Sun et al.[21] propose a weakly supervised framework for learning and recognizing complex user intents from multi-app activity, motivated by examples such as planning dinner with friends across restaurant, messaging, and calendar apps. The framework builds an inventory of intents from a small set of task-oriented user utterances and then recognizes intents from app sequences using graph-based semi-supervised learning, with additional components such as app2vec representations and sequence labeling to extract "content apps" from noisy sequences.

Scenario construction takeaways (Cluster 3). These works treat *language and interaction traces as cue channels*, including utterance-level intents and entities, turn structure and fallback behaviors (confidence thresholds, rephrase requests), cross-lingual handling via translation, and weakly supervised discovery of higher-level intents from interaction sequences.

Importantly, task-intent labels in these systems (for example, "book a table" or "find a restaurant") are not the same object as observer-attributed narratives of social intention (for example, "avoiding conflict" or "trying to impress"). However, they provide techniques for extracting structured signals that can serve as inputs to internal and external script reasoning about restaurant interactions.

3.4 Synthesis: dimensions of variation for scenario generation

The takeaways above can be organized as a compact set of scenario variables aligned with the 3Cs distinction that cues do not, by themselves, determine psychological meaning[17]. Script theory further motivates separating external-script structure from internal-script configuration, since internal scripts are shaped by goals and perceived situational characteristics[6].

A. Cues (objective, potentially observable). Persons and roles present; activities (enter, sit, order, wait, eat, pay, leave); objects and object states (plates, glasses, bill); events (service failure, norm violation); location/space and time (rush vs. quiet, waiting duration).

B. External-script (normative structure). Play type (routine dining, service recovery, joint decision making, payment coordination); scene sequencing (enter $\rightarrow$ seat $\rightarrow$ order $\rightarrow$ wait $\rightarrow$ eat $\rightarrow$ pay $\rightarrow$ leave, with disruptions such as failures or violations); role expectations (customer, server, manager, robot) and responsibility attributions that change normative interpretation.

C. Internal-script (perspectival configuration). Observer goals (for example, norm-enforcing vs. conflict-avoidant) and perceived situational characteristics (the inferred meaning layer derived from cues) that guide which scriptlets and roles are considered plausible.

D. Divergence points. Points where the same cue constellation plausibly maps to different internal scriptlets, role assignments, or trait inferences, as illustrated by norm violations (food theft), responsibility shifts (robot vs. customer error), preference under-specification ("no preference"), and incentive rule changes (split vs. individual pay).

4 Results (Scenario Creation)

This section introduces two short, interaction-focused scenarios derived from the literature survey. Each scenario is presented with three *variations* (different versions of the same interaction) that manipulate a small set of variables to change open-endedness. The design follows the 3Cs separation of cues, characteristics, and classes, where cues are observable but do not uniquely determine psychological meaning[17]. It also follows script theory's distinction between external scripts (normative plays, scenes, roles, scriptlets) and internal scripts that are dynamically configured based on goals and perceived situational characteristics[6].

4.1 Scenario 1 (expands existing literature): Robot service recovery at the table

Motivation. This scenario expands service-recovery findings on the interaction between robot social perception (warm vs competent) and recovery style (humorous vs rational), including responsibility attribution as a boundary condition[19]. The goal is not to replicate the original design, but to turn the same ingredients into a compact vignette that supports multiple plausible observer narratives under different cue visibility and norm clarity.

Base interaction (shared across variations)
A service robot delivers a drink to a table. A small mistake occurs (spill, wrong item, or delayed service). The robot performs a brief recovery utterance and action (apology, replacement, escalation). A customer reacts (facial expression, gesture, posture), and a companion at the table may also react.

Variation 1A: Clear responsibility, low ambiguity (lower openendedness).

- **Class.** Service recovery.
- **Cues.** Robot moves arm toward table, cup tilts, liquid visibly spills onto napkin. Robot immediately pauses, retrieves replacement from tray. Customer flinches then relaxes.
- **External script (normative).** Service failure $\rightarrow$ apology $\rightarrow$ prompt correction. Roles: robot as server, customer as recipient, companion as bystander. Scriptlets: acknowledge fault, fix quickly, restore comfort.
- **Internal script variants (at least 2).**
  - Customer-internal A: "accident, no harm," goal is smooth interaction, interprets robot as trying to help.
  - Customer-internal B: "incompetence," goal is norm enforcement, interprets robot as careless, expects escalation to human staff.
- **Divergence points.** Customer facial response (smile vs frown), distance moved away from spill, whether the robot's pause is read as "careful" or "confused."
- **Allowed observation.** Video-only (no audio), activity and object-state cues (spill, replacement).

Variation 1B: Response style mismatch (medium openendedness). Same physical failure as 1A, but the robot uses a *humorous* recovery line while displaying a competent-looking execution (fast cleanup). This targets the humor vs rational manipulation and the warmth vs competence interpretation tension.

- **Internal script variants.**
  - Customer-internal A: humor signals warmth, goal is tension reduction, interprets robot as considerate.
  - Customer-internal B: humor violates "serious apology" scriptlet, goal is respect, interprets humor as dismissive.

- **Divergence points.** Whether the customer's brief laugh is read as genuine relief or as awkward compliance (politeness).
- **Allowed observation.** Two conditions are possible without changing the story: (i) video-only, or (ii) video plus *robot-generated text* (the recovery utterance is available as a log, even if human audio is not). This keeps the interaction identical while changing the cue layer.

Variation 1C: Ambiguous responsibility (higher openendedness). A spill happens, but the camera does *not* clearly show whether the robot caused it (occlusion by the customer's arm). The customer says or gestures something toward the robot. The robot responds with either humor or rationality, but observers lack clear causal cues.

- **Cues.** Occlusion at the moment of spill, mixed gaze shifts between cup and robot, companion leans in as if to help.
- **External script.** Competing scripts are plausible: "robot error" vs "customer bumped cup." Locus of responsibility becomes a latent variable that shapes interpretation.
- **Internal script variants.**
  - Observer-internal A: assumes robots are reliable, attributes fault to customer, reads robot response as patient.
  - Observer-internal B: assumes robots are error-prone, attributes fault to robot, reads robot response as excuse-making.
- **Divergence points.** Causal attribution under partial observability, plus whether humor is read as affiliative or as blame-shifting.
- **Allowed observation.** Video-only; optional object-state tracking (cup, napkin, spill) but no audio.

How to tune the inner workings and personalize (Scenario 1).

- **Increase or decrease openendedness by cue control:** add occlusion, reduce camera angle coverage, or remove the moment-of-cause cue (as in 1C).
- **Change external-script strength:** make the failure minor (late water refill) vs salient (spill on clothing), or add a manager role who appears immediately (stronger script) vs not (weaker script).
- **Personalization variable:** relationship between diners (date vs colleagues), cultural expectations about apology style, prior robot familiarity (first time vs frequent), and the customer's goal framing (conflict-avoidant vs norm-enforcing).
- **Narrative targets to elicit from annotators/systems:** perceived warmth/competence of robot, sincerity/relief, and inferred customer intention (de-escalate, demand accountability, save face).

4.2 Scenario 2 (fills a gap): "No preference" during the pay decision

Motivation. Prior work shows that "no preference" communication in joint decisions leads recipients to infer undisclosed preferences, increasing decision difficulty and reducing liking and enjoyment[12]. Separately, payment rules (even split vs individual pay) change incentives and can increase consumption under splitting[7]. A gap for scenario-based intention research is a compact restaurant interaction that combines (i) a strong external script (paying norms) with (ii) a communication move ("no preference") that systematically increases interpretive divergence, yielding multiple plausible observer narratives of social intention.

Base interaction (shared across variations)

A server places the bill (or payment terminal) on the table and asks how the group wants to pay. One diner (Person A) responds with a "no preference" style answer. The other diner (Person B) hesitates and looks at A, then at the bill, then back at the server.

Variation 2A: Simple "no preference," minimal asymmetry (lower openendedness).

- **Class.** Payment coordination, joint decision.
- **Cues.** Bill placed on table, two diners seated. Person A leans back, neutral face, small shrug. Person B scans bill and pauses.
- **External script (normative).** Scene: paying. Scriptlets: decide split vs together, avoid awkwardness, be fair. Roles: payer(s), companion(s), server.
- **Internal script variants (at least 2).**
  - A-internal A: conflict avoidance, goal is smooth social interaction, truly indifferent.
  - A-internal B: preference withholding, goal is image management, wants the other person to choose.
- **Divergence points.** Person B's inference of "undisclosed preference" vs "true indifference," and whether B interprets A as cooperative or burdensome.
- **Allowed observation.** Video-only plus object cues (bill, payment terminal). Optional: if the question is typed into a kiosk, the "no preference" phrase can be available as text without recording audio.

Variation 2B: Visible consumption asymmetry (medium openendedness). Same as 2A, but cues suggest asymmetry (A had a visibly larger or more expensive item, or an extra drink is on A's side). This links the paying decision to incentive and fairness concerns consistent with split-bill dynamics.

- **Internal script variants.**
  - B-internal A: suspects strategic exploitation, goal is fairness enforcement, reads A's "no preference" as evasive.
  - B-internal B: assumes social norm of splitting among friends, goal is harmony, reads A as easy-going.

- **Divergence points.** Whether the visible asymmetry is interpreted as intentional advantage-taking, ignorance, or justified (special occasion).
- **Allowed observation.** Video-only (objects on table and bill), no audio. This forces narratives to be constructed from cues that do not uniquely determine meaning.

Variation 2C: Competing external scripts (higher openendedness). Same interaction, but a third cue suggests a different normative play: a company card on the table, a birthday candle, or a "treating" gesture (A reaches toward wallet then stops). Now multiple external scripts are plausible (split, treat, reimburse), increasing open-endedness without changing the core scene.

- **External script (normative).** Competing plays: "friends split," "one person treats," "business expense." The question "how to pay" is under-specified relative to the cues.
- **Internal script variants.**
  - A-internal A: intends to treat but wants B to accept without discomfort, so uses "no preference" to let B choose.
  - A-internal B: intends to avoid paying more, relies on ambiguity, uses "no preference" to shift agency to B.
  - B-internal A: interprets A's move as politeness and tries to refuse treatment.
  - B-internal B: interprets as social pressure and feels trapped.
- **Divergence points.** Role confusion (payer vs treat-giver), "no preference" as generosity vs avoidance, and B's decision difficulty escalation.
- **Allowed observation.** Video-only plus object cues. Optional: payment terminal UI logs (split selected, amounts) as non-audio cues.

How to tune the inner workings and personalize (Scenario 2).

- **Openendedness variables (two dimensions):** (i) external-script clarity (clear split norm vs competing treat-business norms), and (ii) cue completeness (visible asymmetry, occlusion of bill totals, presence or absence of "treating" artifacts).
- **Personalization variables:** relationship type (close friends, first date, coworkers), cultural expectations around splitting, and budget sensitivity (student vs professional). These shift which internal scripts are plausible and which characteristics are inferred.
- **Language as optional cue channel:** the "no preference" phrase can be delivered as speech (not recorded) or via text on a kiosk or chat interface, allowing the same story to be tested with different allowed-observation policies.
- **Narrative targets to elicit:** inferred undisclosed preference, perceived decision difficulty, inferred intentions (avoid conflict, manage impressions, exploit ambiguity, signal generosity), and perceived fairness norms.

5 Responsible Research

This project studies perceived social intention in restaurant scenarios, which raises ethical concerns because social-intention inferences can be sensitive, subjective, and potentially stigmatizing if treated as objective truth. To reduce privacy risks, the scenarios and proposed sensing assumptions focus on *apparent* or *perceived* intention and on privacy-constrained cues (e.g., activity and object-state cues) rather than identifying individuals or recording rich personal data. The work also treats disagreement as informative variation rather than annotation noise, which helps avoid enforcing a single "correct" interpretation of people's behavior.

Use of generative AI (ChatGPT). ChatGPT was used as a writing and analysis assistant in three ways: (i) to support analysis and synthesis of the selected papers when drafting the literature survey, (ii) to support paper discovery during the snowballing step by suggesting potentially relevant citations based on already selected works, and (iii) to help rewrite and rephrase parts of the manuscript (e.g., improving clarity, structure, and LaTeX consistency). All outputs were reviewed by the author, and factual claims were checked against the original sources before inclusion.

Reproducibility. The literature study is designed to be reproducible: the Scopus query used to retrieve candidate papers is provided in an appendix, and the screening procedure and paper selection counts are reported. Scenario construction is also made reproducible through a fixed scenario template and explicitly documented variables (cue-layer, external-script, and internal-script variations), enabling other researchers to recreate or extend scenarios while controlling openendedness. Limitations remain because scenario plausibility and narrative diversity ultimately depend on human interpretation, so future work should include standardized annotation protocols and reporting practices to support comparability across studies.

6 Discussion

The results are two short restaurant interaction case studies with three variations each (Section 4). They are derived from the literature survey (Section 3) and structured using the cues-characteristics-classes bridge and internal vs. external scripts framing.

Comparison to prior work. Scenario 1 (robot service recovery) directly extends Shan et al.[19] work by turning its key factors (warm vs. competent robot, humorous vs. rational recovery, and locus of responsibility) into explicit openendedness variables. In particular, Variation 1C introduces partial observability and unclear responsibility, which is not the focus of the original controlled experiments, but is plausible in real restaurant sensing settings.

Scenario 2 ("no preference" during paying) combines two lines of evidence that are separate in prior work: Kim et al.[12] (recipient inferences of hidden preferences and increased decision difficulty) and Gneezy et al.[7] (norm and incentive tensions around split versus individual pay). The scenario variations treat these as interacting scripts: the same "no preference" cue can be read as conflict avoidance, strategic ambiguity, or generosity signaling depending on perceived norms and visible asymmetries.

Reflection and explanation. A key conclusion from the construction process is that openendedness can be created systematically by separating observable cues from inferred meanings, as argued in Rauthmann and Sherman[17] (cues do not uniquely determine psychological characteristics). Script theory further explains why the same cues can support multiple narratives: internal scripts are configured based on observer goals and perceived situational characteristics rather than being fixed by the external script alone, as described in Fischer et al.[6].

Gap in knowledge. Across the reviewed restaurant and restaurant-like works, there is still no standard way to (i) annotate or represent *multiple plausible* social-intention narratives for the same cue constellation (most works assume a single label or outcome), and (ii) evaluate systems on narrative diversity and plausibility rather than single-label accuracy.

7 Conclusions and Future Work

This work addressed the question: *How to create scenarios in a restaurant setting that allow investigation into how humans and intelligent systems construct multiple plausible narratives of social intention?* We conclude that scenario creation can be made systematic by (i) separating what is **observable** from what is **inferred**, and (ii) explicitly modeling how restaurant interactions are structured by both **external scripts** (normative plays, scenes, roles, and expected scriptlets) and **internal scripts** (observer-dependent interpretations shaped by goals and perceived situational characteristics). The paper puts this into a reusable scenario template and a set of variables that control openendedness by varying cue completeness, norm clarity, and observer perspective without changing the underlying situation class.

The main contribution is therefore methodological: a literature-grounded pipeline that turns prior findings into scenario ingredients (cue sets, narrative targets, and divergence triggers) and then instantiates them as short, interaction-focused case studies with controlled variations. The resulting scenarios illustrate two complementary routes to openendedness: (1) creating ambiguity through partial observability and responsibility attribution in service recovery, and (2) creating ambiguity through norm competition and strategic underspecification in joint payment decisions. Together, they show how to design restaurant interactions where the same cue constellation plausibly supports multiple intention narratives, which is essential if intelligent systems are to be evaluated beyond single-label "ground truth" intention prediction.

Several open issues remain. First, there is no standard method to **represent and evaluate** sets of plausible narratives (diversity, plausibility, and explanation quality) rather than a single correct label. Second, it is unclear how to **collect annotations** that reflect internal scripts without forcing agreement or collapsing disagreement into noise. Third, privacy constraints raise practical questions about which cue channels are sufficient for narrative generation and how missing cues change narrative plausibility.

Future work should therefore prioritize: (i) human studies that validate whether the proposed variations reliably create different levels of openendedness and elicit multiple stable narrative clusters, (ii) development of evaluation metrics for multi-narrative outputs (for example, measuring coverage and plausibility rather than accuracy to a single label), (iii) systematic expansion of the scenario set across additional restaurant plays (ordering conflicts, queueing, seating, staff-customer misunderstandings) while keeping the same template for comparability, and (iv) benchmarking intelligent systems on producing and justifying multiple narratives under controlled cue access.

References

[1] Thomas R. Alley. Effects of Minor Food Theft on Social Perceptions of Culprits and Victims. *Evolutionary Psychological Science*, 4(1):90–97, March 2018.

[2] Diego Bufquin, Robin DiPietro, Marissa Orlowski, and Charles Partlow. Social evaluations of restaurant managers: The effects on frontline employees' job attitudes and turnover intentions. *International Journal of Contemporary Hospitality Management*, 30(3):1827–1844, March 2018.

[3] Diego Bufquin, Robin DiPietro, Jeong-Yeol Park, and Charles Partlow. Effects of Social Perceptions and Organizational Commitment on Restaurant Performance. *Journal of Hospitality Marketing & Management*, 26(7):752–769, October 2017.

[4] Diego Bufquin, Robin B. DiPietro, Charles Partlow, and Scott J. Smith. Differences in social evaluations and their effects on employee job attitudes and turnover intentions in a restaurant setting. *Journal of Human Resources in Hospitality & Tourism*, 17(3):375–396, July 2018.

[5] Lkd Fernando and Gu Ganegoda. ResBot: A Bi-Lingual Restaurant Booking Conversational Artificial Intelligence. In *2023 8th International Conference on Information Technology Research (ICITR)*, pages 1–6, December 2023. ISSN: 2831-3399.

[6] Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist*, 48(1):56–66, January 2013. _eprint: https://doi.org/10.1080/00461520.2012.748005.

[7] Uri Gneezy, Ernan Haruvy, and Hadas Yafe. The inefficiency of splitting the bill. *The Economic Journal*, 114(495):265–280, 2004. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2004.00209.x.

[8] Hayley Hung, Litian Li, Jord Molhoek, and Jing Zhou. The Discontent with Intent Estimation In-the-Wild: The Case for Unrealized Intentions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, Honolulu HI USA, May 2024. ACM.

[9] Rajmohan Kadavil and M. Usha. Exploring the influence of sensory brand recognition on brand loyalty within the quick service restaurant industry: an analysis of branded food retailers in Kerala. *Salud, Ciencia y Tecnología - Serie de Conferencias*, 3:899–899, January 2024.

[10] Moritz Kaiser, Dejan Arsic, Benedikt Hornler, Martin Hofmann, and Gerhard Rigoll. THE NOLDUS DATABASE: AUTOMATED RECOGNITION OF RESTAURANT RELATED ACTIVITIES FOR THE RESTAURANT OF THE FUTURE. *International Workshop on Image Analysis for Multimedia Interactive Services*.

[11] Dongkeun Kim, Youngkil Song, Minsu Cho, and Suha Kwak. Towards More Practical Group Activity Detection: A New Benchmark and Model, July 2024. arXiv:2312.02878 [cs].

[12] Nicole You Jeung Kim, Yonat Zwebner, Alixandra Barasch, and Rom Y. Schrift. You Must Have a Preference: The Impact of No-Preference Communication on Joint Decision Making. *Journal of Marketing Research*, 60(1):52–71, February 2023.

[13] Paige N. Michener, Joanna Cassella, and Todd R. Schachtman. Retrospective revaluation effects during interpersonal attributions. *Learning and Motivation*, 87:101995, August 2024.

[14] Vahid Mokhtari, Luis Seabra Lopes, and Armando J. Pinho. Experience-Based Robot Task Learning and Planning with Goal Inference. *Proceedings of the International Conference on Automated Planning and Scheduling*, 26:509–517, March 2016.

[15] Alexandros Papangelis, Karthik Gopalakrishnan, Aishwarya Padmakumar, Seokhwan Kim, Gokhan Tur, and Dilek Hakkani-Tur. Generative Conversational Networks. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 111–120, Singapore and Online, 2021. Association for Computational Linguistics.

[16] Marco Quiroz, Raquel Patiño, José Diaz-Amado, Yudith Cardinale, Marco Quiroz, Raquel Patiño, José Diaz-Amado, and Yudith Cardinale. Group Emotion Detection Based on Social Robot Perception. *Sensors*, 22(10), May 2022.

[17] John F. Rauthmann and Ryne A. Sherman. Conceptualizing and measuring the psychological situation. In *Measuring and Modeling Persons and Situations*, pages 427–463. Elsevier, 2021.

[18] Kisang Ryu and Jin-Soo Lee. Examination of Restaurant Quality, Relationship Benefits, and Customer Reciprocity From the Perspective of Relationship Marketing Investments. *Journal of Hospitality & Tourism Research*, 41(1):66–92, January 2017.

[19] Minghui Shan, Zhenzhong Zhu, Haipeng (Allan) Chen, and Sijie Sun. Service robot's responses in service recovery and service evaluation: the moderating role of robots' social perception. *Journal of Hospitality Marketing & Management*, 33(2):145–168, February 2024. _eprint: https://doi.org/10.1080/19368623.2023.2246456.

[20] Nadhifa Sofia, Edi Winarko, and Sigit Priyanta. Incorporating IndoBERT into DIET Classifier to Enhance Intent and Entity Recognition in Culinary Chatbot, 2025.

[21] Ming Sun, Aasish Pappu, Yun-Nung Chen, and Alexander I. Rudnicky. Weakly supervised user intent detection for multi-domain dialogues. In *2016 IEEE Spoken Language Technology Workshop (SLT)*, pages 91–97, December 2016.

[22] Ada V. Taylor, Roman Kaufman, Michael Huang, and Henny Admoni. Group Activity Recognition in Restaurants to Address Underlying Needs: A Case Study. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 236–243, August 2022. ISSN: 1944-9437.

[23] Ahmet ŞahiNoğlu and Fatma Başar. The Impact of Restaurant Experience on Customer Satisfaction and Behavioral Intention: The Case of Van. *Journal of Academic Tourism Studies*, 4(Cilt: 4 Sayı: Özel Sayı):63–76, 2023.

A Scopus Search Query

```
(
  TITLE-ABS-KEY(”social intention*”) OR
  TITLE-ABS-KEY(”intent* recognition”) OR
  TITLE-ABS-KEY(”intent* estimat*”) OR
  TITLE-ABS-KEY(”intent detection”) OR
  TITLE-ABS-KEY(”intent* predict*”) OR
  TITLE-ABS-KEY(”behavio*r interpret*”) OR
  TITLE-ABS-KEY(”action understanding”) OR
  TITLE-ABS-KEY(”social percept*”) OR
  TITLE-ABS-KEY(”action prediction”) OR
  TITLE-ABS-KEY(”goal inference”) OR
  TITLE-ABS-KEY(”narrative expl*”) OR
  TITLE-ABS-KEY(”plausible narrative*”) OR
  TITLE-ABS-KEY(”intent* inferenc*”) OR
  TITLE-ABS-KEY(”goal recogn*”) OR
  TITLE-ABS-KEY(”goal estimat*”) OR
  TITLE-ABS-KEY(”goal inferenc*”) OR
  TITLE-ABS-KEY(”behavio*r recogn*”) OR
  TITLE-ABS-KEY(”behavio*r estimat*”) OR
  TITLE-ABS-KEY(”behavio*r inferenc*”) OR
  TITLE-ABS-KEY(”behavio*r understand*”)
  )
  AND
  (
  TITLE-ABS-KEY(”restaurant*”) OR
  TITLE-ABS-KEY(”diner*”) OR
  TITLE-ABS-KEY(”dining”) OR
  TITLE-ABS-KEY(”eatery”) OR
  TITLE-ABS-KEY(”tavern*”) OR
  TITLE-ABS-KEY(”cafe*”) OR
  TITLE-ABS-KEY(”café*”) OR
  TITLE-ABS-KEY(”pizzeria*”) OR
  TITLE-ABS-KEY(”canteen*”) OR
  TITLE-ABS-KEY(”drive-in*”) OR
  TITLE-ABS-KEY(”doughtnut shop*”) OR
  TITLE-ABS-KEY(”hamburger stand*”) OR
  TITLE-ABS-KEY(”hotdog stand*”) OR
  TITLE-ABS-KEY(”waiter*”) OR
  TITLE-ABS-KEY(”waitstaff*”) OR
  TITLE-ABS-KEY(”food service*”) OR
  TITLE-ABS-KEY(”foodservice*”) OR
  TITLE-ABS-KEY(”food outlet*”) OR
  TITLE-ABS-KEY(”food establishment*”) OR
  TITLE-ABS-KEY(”catering”)
  )
  AND NOT (
  TITLE-ABS-KEY(”social media”) OR
  TITLE-ABS-KEY(”online”) OR
  TITLE-ABS-KEY(”twitter”) OR
  TITLE-ABS-KEY(”facebook”) OR
  TITLE-ABS-KEY(”animal*”) OR
  TITLE-ABS-KEY(”husbandry”)
  )
```