Social Repair

Alafate Abulimiti[*] and **Vladislav Maraev**[†,*]

[*]ISIR - Institut des Systèmes Intelligents et de Robotique
Sorbonne Unversity
[†]Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
Correspondence: alafate.abulimiti@isir.upmc.fr

Abstract

Social repair encompasses the crucial strategies individuals employ to maintain and restore social relationships and interpersonal harmony when interactions go awry. While existing research on repair has often prioritized communicative and informational aspects, the social and relational dimensions have received less focused attention. This paper addresses this gap by developing a formal representation of social repair centered on the dynamic coordination of social beliefs between interlocutors. Drawing insights from social theories of interaction, Type Theory with Records within Dialogue Game Board (DGB) theory, we define social repair and specify its key components, including an agent's social beliefs. We then delineate the interactional dynamics through which social misalignments are perceived and addressed. These dynamics can lead to distinct reactive strategies by an interlocutor who recognizes a potential issue: accommodation to the other's social belief, an attempt to calibrate the other's interpretation, or ignoring the perceived misstep. This framework provides a systematic approach to understanding and modeling the nuanced processes of social repair, offering important implications for developing more socially aware agents.

1 Introduction

Effective communication and social interaction are essential for building and maintaining relationships (Duck, 1994), achieving common goals (Pentland, 2012), and navigating complex social situations (Forgas and Jones, 1985). However, even with the best intentions, individuals may encounter moments where their actions, words, or behaviors are misaligned with the social expectations or beliefs of others (Goffman, 1967). In such cases, social repair becomes crucial for restoring harmony, preventing further damage to the relationship, and ensuring that the interaction remains productive and "connective".

This paper investigates the concept of *social repair*, which encompasses a range of actions that individuals use to address and resolve social missteps or conflicts. Despite the significant advances in our understanding of repair, much of the existing research has focused primarily on the communicative and informational aspects of repair, with less attention paid to its social and relational dimensions. While some studies have touched on the role of politeness and face management (Brown and Levinson, 1987; Domenici and Littlejohn, 2006), the broader social functions and implications of repair have often been overlooked. This is where the concept of social repair comes in. As introduced by Abulimiti et al. (2021), social repair refers to the processes and strategies that individuals use to maintain and restore social relationships and interpersonal harmony in the face of interactional troubles or breaches. While their work first identified and named this crucial concept, the present paper builds upon this foundation by proposing a formal model that specifies the underlying cognitive and interactional mechanisms of social repair. Specifically, our work models social repair as a reactive phenomenon, triggered by a perceived problematic utterance or a demonstrated incompatibility of social beliefs. Through this formalization, we aim to bridge the gap between the study of communicative repair and the broader social theories of interaction.

To illustrate the importance of social repair, consider the following constructed example of a doctor-patient interaction:

DOCTOR: Good morning, Mrs. Johnson. How have you been feeling since your last visit?

PATIENT: To be honest, Doctor, I've been feeling worse. The medication you prescribed doesn't seem to be helping.

DOCTOR: [*frowning*] That's not good. Are you sure you've been taking the medication

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

as directed?

PATIENT: [*defensively*] Of course, I have! Doctor! I know how to follow instructions.

DOCTOR: [*realizing the misstep*] I didn't mean that... Let's take a closer look at your symptoms and see if we can find a better solution together.

PATIENT: [*relaxing slightly*] Thank you, Doctor.

In this example, the doctor's initial response to the patient's concerns about the medication's effectiveness comes across as accusatory, suggesting that the patient may not be following the prescribed instructions. The patient, feeling defensive and disrespected, responds with a sharp comment. Recognizing the potential damage to the doctor-patient relationship and the need to maintain trust and open communication, the doctor engages in *social repair*. By apologizing for the unintended implication, acknowledging the patient's concerns, and proposing a collaborative approach to finding a solution, the doctor demonstrates their commitment to repairing the social misstep and maintaining a positive, productive interaction. By understanding and applying social repair, individuals can effectively address misunderstandings, mitigate conflicts, and foster stronger, more resilient relationships.

2 Related Work

2.1 Communicative Repair

Communicative repair is defined that participants to take the actions in a conversation to identify and resolve problems in understanding or communication breakdowns. Schegloff et al. (1977) first identified the basic mechanisms and patterns of repair in talk-in-interaction, distinguishing between self-initiated and other-initiated repair, as well as self-repair and other-repair. Building on this foundational work, researchers have explored the various forms and functions of repair in different contexts and settings (Levelt, 1983; Clark and Schaefer, 1989). More recently, the study of repair has been further advanced by the work of Ginzburg (2012) and colleagues, who have developed a comprehensive framework for modeling dialogue and interaction, known as the KoS (Ginzburg, 2012). This framework provides a formal account of the interactional dynamics of repair, including the role of clarification questions and feedback in resolving misunderstandings (Ginzburg et al., 2003; Purver et al., 2018).

Dingemanse and Enfield (2024) highlight that communicative repair, crucial for information robustness, also organizes social accountability, forming a foundation for human language.

Researchers such as Traum (1994) and Heeman and Allen (1999) developed computational models of repair that aimed to capture the complex interactional dynamics of human conversation. These models were based on the idea that repair is a fundamental mechanism for maintaining common ground and ensuring mutual understanding between interlocutors. More recently, researchers have explored the role of non-verbal cues and embodied signals in coordinating repair and maintaining mutual understanding in human-robot interaction (Gross et al., 2017). Overall, while these and other computational models have addressed various facets of communication repair, dedicated computational models specifically for social repair, as conceptualized in this work (i.e., focusing on the alignment of social beliefs and relational maintenance), appear to be less prevalent or represent an important avenue for future research.

2.2 Social Theories of Interaction

The study of social repair in dialogue and interaction is deeply rooted in various social theories that attempt to explain the dynamics of human communication and relationships. One of the most influential social theories related to the study of social repair is Goffman's (1959) study of face and self-presentation in everyday life. Goffman argues that individuals apply different strategies to preserve their face and manage the impressions they convey to others, which may involve various forms of remedial work. This inherent belief is formed prior to the interaction and is referred to as "Front".

Building on Goffman's insights, Brown and Levinson (1987) proposed a comprehensive theory of politeness that explains how individuals use language to manage face. Their framework has been widely influential in the study of social interaction and has informed many studies on the role of politeness in repair and miscommunication (e.g., Domenici and Littlejohn, 2006; Oetzel et al., 2001).

Another one of the key theoretical perspectives that have influenced the study of social repair is the concept of rapport, which was developed by Tickle-Degnen and Rosenthal (1990); Spencer-Oatey (2005) and Zhao et al. (2014). Rapport is defined as a positive, harmonious and smooth feeling

between interlocutors. Maintaining and restoring rapport is one of the key functions of social repair in interactions, as it helps to ensure the overall quality and effectiveness of ongoing interactions.

The study of social repair also draws on the theories of empathy and perspective taking, which emphasize the importance of understanding and responding to the thoughts, feelings and needs of others in social interactions. As noted by Decety and Jackson (2004) and others (Davis, 1983; Hoffman, 1996; Eisenberg, 2000), empathy plays a crucial role in facilitating effective communication, cooperation and conflict resolution, all of which are central to the process of social repair. And this proactive behavior may be effective in triggering repair before conflicts (or the demonstration of conflicts). We are referred to in section 3 as social self-repair.

3 Formal Representation

3.1 Defining Social Repair and Social Beliefs

Social repair involves maintaining and restoring a social relationship. Fundamentally, social repair is a dynamic process of coordinating social beliefs. Drawing an analogy to Clark's concept of common ground (Clark and Brennan, 1991; Clark, 1996), where interactants give each other evidence that their contributions are understood sufficiently for current purposes, social repair involves interactants working towards a mutual recognition that their social beliefs about the interaction and each other are aligned, or are being actively managed towards alignment, to a degree that sustains their ongoing relationship and interactional goals.

Improved social relations are often the result of achieving and maintaining this coordinated understanding of social beliefs. Specifically, in its core mechanism, social repair is the process by which an individual, referred to as A, deliberately attempts to align their beliefs or perceptions with those of another individual, B, regarding their interaction. This alignment process typically occurs after A has presented their beliefs (or the clues of beliefs) in a manner that is inconsistent or misaligned with B's beliefs. In this context, the term "beliefs" specifically refers to those pertaining to the social dynamics of the interaction (i.e., "social beliefs"), as opposed to a broader range of beliefs. For example, in a dyadic interaction, A believes that B thinks A is friendly. While this core mechanism is often illustrated with two participants (A and B) for clarity, the principles of social repair extend to multi-party dialogues where the alignment of social beliefs is negotiated among several participants. Moreover, the presence of overhearers (Traum, 2003) (or even the awareness that an interaction is being recorded) can significantly influence the dynamics and goals of social repair, for instance, by heightening the importance of face-saving maneuvers or altering the perceived need for explicit alignment depending on the audience.

Social repair can be considered a specific type of content repair, which addresses problems or misunderstandings in the content or meaning of a socially interpreted utterance (Schegloff et al., 1977; Schegloff, 1992). While content repair focuses on the informational content of an utterance, social repair deals with the social beliefs and interpretations of the interactants within a dialogical context.

3.2 Theoretical Foundations

A robust formalization of social repair necessitates grounding in established theories of agency, information representation, and dialogue dynamics. The core of social repair lies in the coordination of *social beliefs* concerning the social interpretation of a dialogue event.

Our model posits that social repair is triggered by social belief incompatibility, a choice grounded in established cognitive frameworks (Georgeff and Rao, 1991). While social expectations are modeled here as beliefs, constructs like goals and values can potentially be modeled in a similar fashion.

To capture the dialogue context with the necessary precision and structural integrity, our formal model uses Type Theory with Records (Cooper and Ginzburg, 2015) along with Ginzburg's Dialogue Game Board (DGB) theory (Ginzburg, 2012). This framework offers a powerful and flexible means to define the rich, feature-based information structures that constitute an agent's cognitive state (such as a private knowledge of an individual encompassing social beliefs) and the shared interactional context. The use of typed feature structures ensures that our representations are well-defined for a formal analysis, which is essential for developing a computationally explicit theory of social repair.

The total information state (see the representation 1) contains the shared information that can be derived from publicized interaction and also private information that includes private beliefs, plans. The shared information state, at least, monitors the shared assumptions, dialogue content said by

the interlocutors, and also the focus of the current dialogue interaction (i.e., Questions Under Discussion).

3.3 Social Interpretation Function

The notion that individuals engage in "self-interpretation of the observed behaviors" of others is strongly supported by research in social cognition, with Theory of Mind (ToM) being a particularly crucial cognitive capacity. ToM refers to the ability to attribute mental states—such as beliefs, desires, intentions, emotions, and knowledge—to oneself and to others, and to understand that these mental states can differ and are what often drive behavior (Wellman, 2018; Call and Tomasello, 2008; Van Overwalle, 2009). In essence, when we observe someone's actions or expressions, we are not merely registering the physical movements; we are actively trying to infer the underlying mental state that produced them. This inference *is* an act of interpretation.

Social cognition, as a broader field, encompasses ToM and investigates the cognitive processes we use to understand, store, and apply information about other people and social situations (Van Overwalle, 2009). It fundamentally involves making social inferences about these implicit dialogue states, which are not directly observable but are deduced from a pattern of verbal and nonverbal cues (Adolphs, 2009; Van Overwalle, 2009). This means that during face-to-face interactions, individuals are constantly engaged in a process of decoding and making sense of the myriad social signals being emitted by their interaction partners—from facial expressions and vocal prosody to body language and gaze (Hamilton and Holler, 2023; Vinciarelli et al., 2009; Ekman and Rosenberg, 1997).

This interpretive mechanism is not a passive, one-off event but a continuous and dynamic process. As individuals interpret the behavior of others, these interpretations directly inform and shape their own subsequent responses (Redcay et al., 2010; Hömke et al., 2025). For example, perceiving an eyebrow furrow as a sign of confusion (an interpretation of a mental state) might lead a speaker to elaborate on their point (a responsive behavior) (Hömke et al., 2025). This creates a feedback loop where one person's interpreted behavior and subsequent response become a new social signal for the other person to interpret, highlighting the inherently interactive and reciprocal nature of social understanding (Hamilton and Holler, 2023; Redcay et al., 2010; Jiang et al., 2021).

Therefore, ToM and the broader mechanisms of social cognition provide compelling evidence that humans are equipped with, and constantly utilize, a sophisticated function for interpreting the behaviors of others. This ongoing interpretation allows individuals to predict actions, understand intentions, empathize with emotions, and ultimately, to navigate the complexities of social interactions by responding in a contingent and socially appropriate manner.

3.4 Formal Model of Social Repair

To formally represent the social repair mechanism, we extend the *Private* state of an individual (as described in the representaton 2) to include a set of social beliefs (*SocialBel*) and a repair set (*RepairSet*). We also draw upon Ginzburg's Dialogue Gameboard theory (Ginzburg, 2012), which provides a framework for modeling the interactive nature of dialogue and the grounding of information in a shared context.

We divide the parts of the *SocialBel* into those that the individual has before the interaction, and those that are generated and adapted during the interaction. Therefore, the *SocialBel* consists of the individual's "Front", as defined by Goffman (1959), and containing a set of social interpretations ($\epsilon(p)$) of the facts (p) in the shared context. Here, the social interpretation function maps the observable behaviors and actions of interlocutors to the underlying social meanings and intentions.

The "Front" represents the individual's public persona that one possesses before the interaction, while the social interpretations capture how the individual perceives and interprets the social facts of the interaction. The *RepairSet* contains a set of repair plans (Λ) that the individual can use to address inconsistencies in social beliefs. These repair plans are derived from the individual's past experiences and knowledge of social norms and conventions, as well as the grounding process described in Ginzburg's DGB theory.

(1) a. Total Information State $=_{def}$
$$\begin{bmatrix} \text{dialoguegameboard} : \text{DGBtype} \\ \text{private} : \text{Private} \end{bmatrix}$$

b. DGBType $=_{def}$

spkr: Ind	turn
addr: Ind	owner-
utt-time: Time	ship
c-utt: addressing(spkr, addr, utt-time)	
Facts: Set(Prop)	shared assumptions
VisSit: [InAttention : Ind]	visual field
Pending: list (LocProp)	ungrounded utts
Moves: list (IlocProp)	grounded utts
QUD: poset (Question)	qs under discus
Mood: Appraisal	face

(2) Private $=_{def}$

$$\begin{bmatrix} \text{Agenda:OpenQueue(Action)} \\ \text{Plan: OpenStack(PlanConstruct)} \\ \text{BEL: } \left[\text{SocialBel=}\Phi\text{: Set}(\epsilon(p))\right] \\ \text{Goals: } \begin{bmatrix} \text{GoalsSet: List(Prop)} \\ \text{GoalsIpt = } \Omega\text{: List(Float)} \end{bmatrix} \\ \text{RepairSet: Set(Plan)=}\Lambda \end{bmatrix}$$

3.5 Social Repair Dynamics

The social repair mechanism unfolds through a sequence of interactions and interpretations, which necessitates extending the `Private` state. The process can be characterized as follows:

First, individual A possesses a set of social beliefs ($\phi \in \Phi$), which includes their "Front" and ongoing social interpretations. An interaction occurs where individual B performs an action or makes an utterance, denoted as p_1. A then interprets this action through their social interpretation function, $\epsilon_a(p_1)$. This interpretation by A might reveal an incompatibility (denoted as $\perp\!\!\!\perp$) with one of A's existing social beliefs, ϕ. For instance, p_1 could be perceived by A as a violation of a social norm A upholds, or as an implication of a status dynamic inconsistent with A's "Front". This perceived mismatch is formally $\epsilon_a(p_1) \perp\!\!\!\perp \phi$.

In response to this perceived incompatibility, A produces a subsequent action or utterance, p_2. This response, p_2, serves as an observable *protest* or signal of misalignment from A's perspective[1]. B, in turn, observes p_2 and interprets it using their own social interpretation function, $\epsilon_b(p_2)$. Through this interpretation, B becomes aware that A perceives an issue related to A's social belief ϕ; essentially, B recognizes A's protest and its connection to ϕ, as in $\epsilon_b(p_2) \rightarrow$ (A perceives issue with ϕ due to p_1). The timing and accuracy of B's discovery of A's protest are contingent upon B's interpretive capabilities (ϵ_b) and the broader conversational setting or context.

Once B recognizes A's protest, B faces a critical choice point regarding social repair. B has several strategic options:

1. **Repair by Accommodation (Aligning with A's Social Belief ϕ):** B can select a repair action a_α from their *RepairSet* (Λ) that aims to accommodate A's perspective and restore alignment with A's protested social belief ϕ. This involves B acknowledging the validity of A's concern (implicitly or explicitly) and modifying their stance, clarifying their original intentions behind p_1 in a conciliatory way, or committing to future actions that are compatible with ϕ. For example, if A interpreted B's joke p_1 as disrespectful (where ϕ is an expectation of respect), B might apologize for p_1 or clarify that no disrespect was intended, thereby respecting A's belief ϕ.

2. **Repair by Calibration (Attempting to Shift A's Interpretation or Belief):** Alternatively, B can choose a repair action $a_\alpha \in \Lambda$ that attempts to "calibrate" A's understanding. This means B seeks to modify A's initial interpretation $\epsilon_a(p_1)$ of B's action p_1 (e.g., by explaining, "That wasn't my intention when I did p_1, I meant it as...") or even to negotiate the applicability or substance of A's social belief ϕ in the current context (e.g., "In this informal setting, isn't it usually okay to...? Perhaps ϕ is a bit strict here?"). This approach is more of a negotiation, where B tries to bring A's perspective closer to their own, rather than simply conforming to A's existing belief.

3. **Ignore the Protest:** B may choose, either intentionally (e.g., maintaining power relationship, Guilfoyle, 2003) or unintentionally (e.g., due to misinterpretation or distraction), to disregard A's protest p_2. This course of action typically leaves the social belief incompatibility unresolved and may lead to further interactional trouble or damage to the social relationship.

The chosen repair action a_α is then enacted by B. The success of this repair attempt in resolving the incompatibility and restoring social coordination will depend on A's subsequent interpretation of a_α

[1]The observability of p_2 is crucial, as a non-observable internal *protest* would not be functionally available to B for initiating repair.

and whether it sufficiently addresses the initial perceived misalignment concerning ϕ. This can lead to a new state of mutual understanding, further rounds of repair, or a persistent state of misalignment.

4 Examples

This section illustrates each of the three aforementioned reaction types (accommodation, calibration, and ignoring) with an example.

4.1 Accomodation

The first type of social repair reaction, accommodation, is illustrated by a social interaction commonly observed in France. In this scenario, if an individual directly requests assistance without offering a prior greeting, the request may not be immediately addressed. Instead, the person from whom help is sought might first respond with "Bonjour" and a smile, anticipating a reciprocal greeting. It is typically only after this customary exchange of greetings that the initial request will be politely handled. To provide a concrete illustration of this accommodative dynamic, a constructed example set in a bakery is presented below.

(3) constructed example

A: Deux croissants et une baguette, s'il vous plait. [1]
(two crossitants and one bread, please.)

BAKER: **Bonjour.** [smile] [silence] [2]
(Good morning.)

A: **Bonjour.** [3]
(Good morning.)

BAKER: Voilà, trois euros s'il vous plait. [4]
(Here you are, three euros please.)

In this example, A had a request for food at the bakery first with utterance [1]. However, this direct request is not handled directly by the baker with possible following utterances and actions, as the opposite, baker responds with [2] "Bonjour", a conventional and respectful greeting phrase commonly used in the French society. In this context, where A is the customer and Baker is the service provider, a "Bonjour" is more likely to represent the respect to another interlocutor. Thus, Baker initiated this conventional process with [2] "Bonjour" and waited for mutual greeting with standing in the silence. When A realized this, immediate mutual "Bonjour" is responded, as a consequence, Baker executed the process to handle A's demand with [4].[2]

Now we come closer to the example and analyze this example with our formal model. The interpreted A's initiation [1] $(\epsilon_b(p_1))$[3] is actually incompatible with the Baker's social belief, ϕ might be in baker's social front. This social belief could be: "customer should greet me", and a more elaborative social belief could be: "client should show their respect to the baker in a not-hurry situation". $\epsilon_b(p_1)$ is the baker's interpretation of p_1 which could be: "A asked baker's directly for food without "Bonjour" is a sign of disrespect. In a word, $\epsilon_a(p_1) \perp\!\!\!\perp (\phi \in \Phi)$. As follows, Baker shows their "protest" with [2] "Bonjour"(p_2). A's received baker's utterance [2] (and also silence) and interpreted as $\epsilon_b(p_2)$, this could be: "A's direct demand is impolite or disrespectable". A's [3] "Bonjour" actually socially repairs the situation. As Baker received [3] "Bonjour" and interpreted as a respectable signal where $\epsilon_b(p_3) \mid\mid (\phi \in \Phi)$, the dialogue continues with Baker's action and proper response [4].

Figure 1 visually schematizes this interaction and the underlying social repair process detailed in the bakery example. (1) The customer's initial "Request" [1] on the dialogue layer is interpreted by the baker, leading to an "Incompatibility Demonstration" on the social layer, as the request is misaligned with the baker's social belief. (2) The baker's responsive "Greeting" ("Bonjour" [2]) on the dialogue layer then functions as a 'Protest' on the social layer against this perceived violation of social expectations. (3) Subsequently, the customer's reciprocal "Greeting" ("Bonjour" [3]) on the dialogue layer constitutes the "Accommodation" on the social layer, thereby repairing the social misalignment. (4) This successful repair is acknowledged as the baker proceeds with the "Action" [4] of serving the customer on the dialogue layer, allowing the interaction to "Continue" smoothly on the social layer.

4.2 Calibration

The second type of social repair reaction discussed is calibration. To illustrate this, a real-world example (Example 4) is presented from the British

[2]This example resembles the example given by Dingemanse and Enfield (2024) (cited from Drew, 1997, example 22) as an illustration of using repair to jointly manage social accountability. We thank our anonymous reviewer for bringing this to our attention.

[3]baker's interpretation on p_1

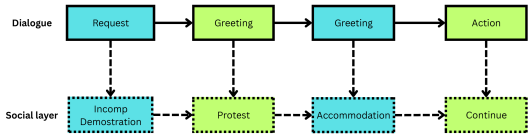

Figure 1: Scheme of "Bonjour" example

National Corpus (BNC). The context for this example is a recorded discussion where the interlocutors are examining potential socio-economic reasons for drug use in deprived areas and the differing perspectives this can elicit.

(4) BNC: J8J (lines 281-294)

ANON 1: Well I just wonder you, there has got to be some kind of <pause> relationship between the fact that most people who take drugs live in really run down deprived areas... You don't think that? ... it's got something to do with the fact that <pause> people have got nothing to do in those areas, and no cha, no prospects, no chance of getting a job and it's actually quite a purposeful way of spending your time [1]

ANON 2: **I think that's really patronizing!** [2]

ANON 1: You're saying that <unclear> No, but **I don't mean to be patronizing.** [3]

ANON 2: if I help with addict that they need to compensate for for <pause> er, things missing in their lives, perhaps they just like it. [4]

ANNO 1: Yes. Yeah! Well maybe. I mean there's nothing wrong with <pause> I mean people do just like. <unclear> [5]

In this example, let Anon 1 be A and Anon 2 be B. The social repair dynamic unfolds as follows:

First, Anon A performs an action, an utterance p_1, in turn [1] by stating their perspective on drug use in deprived areas. Anon B then interprets this p_1 through their social interpretation function, $\epsilon_b(p_1)$. This interpretation by B reveals an incompatibility with one of B's existing social beliefs, ϕ (e.g., a belief that such generalizations are demeaning, an expectation of respectful discourse, or a desire not to be spoken to in a way perceived as patronizing). This perceived mismatch is formally $\epsilon_b(p_1) \perp\!\!\!\perp \phi$. In response to this perceived incompatibility, B produces a subsequent p_2, in turn [2]: "**I think that's really patronizing!**" This response, p_2, serves as an observable protest or signal of misalignment from B's perspective.

A, in turn, observes p_2 and interprets it using their own social interpretation function, $\epsilon_a(p_2)$. Through this interpretation, A becomes aware that B perceives an issue related to B's social belief ϕ; essentially, A recognizes B's protest and its connection to ϕ. A then chooses a repair action $a_\alpha \in \Lambda$ aimed at calibration. This is evident in turn [3] where A responds: "You're saying that <unclear> No, but **I don't mean to be patronizing.**" This repair action a_α attempts to modify B's initial interpretation $\epsilon_b(p_1)$ of A's action p_1. Specifically, A denies the patronizing intent, trying to shift B's understanding of the original statement's meaning or A's underlying intention, rather than immediately accommodating B's interpretation by apologizing for being patronizing. B's subsequent turn [4] ("if I help with addict that they need to compensate for for <pause> er, things missing in their lives, perhaps they just like it.") elaborates on their perspective, suggesting the calibration attempt in turn [3] hasn't fully resolved the issue. The interaction continues with A in turn [5] showing some acknowledgment ("Yes. Yeah! Well maybe."), which might be a further, softer calibration or a move towards accommodation, though the primary calibration move remains the direct denial of patronizing intent in turn [3].

4.3 Ignoring

The "ignore" reaction is demonstrated in the subsequent medical consultation example (Example 5), which is also drawn from the BNC corpus:

(5) BNC: G48 (lines 8-21)

DOCTOR: For who? [1]

PATIENT: Patrick. [2]

DOCTOR: <unclear> getting these prescriptions for? Patrick? **You're daft.** [3]

PATIENT: **It was for Patrick. Er, <pause> see <unclear>** give her something the wee'uns, he's got two wee <pause> [4]

DOCTOR: Has he been bad to that poor wee, wee'un again? [5]

PATIENT: thingummy, abscesses in his er in his tooth. She can nae take him in to the dentists till she gets rid of them. [6]

DOCTOR: <laugh> She's a bad woman. You tell her I said it's just pure neglect. <pause/> Tell her I said that will you? [7]

PATIENT: Mhm. [8]

DOCTOR: And she'll belt you. <pause> Now then <pause> [9]

In this interaction, we consider the Patient as individual A and the Doctor as individual B. The sequence demonstrating the "Ignore the Protest" dynamic unfolds as follows:

The dynamic is initiated when the Doctor (B) makes an utterance, p_1, in turn [1], stating: "<unclear> getting these prescriptions for? Patrick? You're daft." Upon hearing this, the Patient (A) processes this statement through their social interpretation function, $\epsilon_a(p_1)$. This interpretation, particularly the direct assertion "You're daft," likely generates an incompatibility with the patient's (A's) social belief, ϕ—which might involve their self-perception of sound judgment, an expectation of professional respect, or the desire to maintain face. This perceived incongruity is represented as $\epsilon_a(p_1) \perp\!\!\!\perp \phi$. Consequently, the patient (A) offers a response, p_2, in turn [2]: "It was for Patrick. Er, <pause> see <unclear> give her something the wee'uns, he's got two wee <pause>" This utterance, p_2, functions as an observable, though indirect, protest from A's viewpoint, as it seeks to provide a rationale for the action that prompted the Doctor's (B's) critical remark, thereby implicitly challenging the negative evaluation.

Subsequently, the doctor (B) perceives the patient's (A's) utterance p_2 and interprets it via their own social interpretation function, $\epsilon_b(p_2)$. Through this, B would likely register A's attempt at justification and the implicit contestation of the "daft" label, thereby recognizing that A has signaled an issue concerning their social belief ϕ. Despite this, the doctor (B) opts to "Ignore the Protest." This choice is manifested in the doctor's ensuing contributions. In turn [3], the Doctor inquires, "Has he been bad to that poor wee, wee'un again?" and further in turn [4], remarks, "<laugh> She's a bad woman. You tell her I said it's just pure neglect. <pause/> Tell her I said that will you?" These statements from the doctor (B) sidestep any acknowledgment of the patient's (A's) protest regarding the "daft" comment. Rather than addressing A's justification or the challenge to their competence, the doctor redirects the conversation, introducing new lines of criticism aimed at third parties. By failing to engage with the patient's implicit protest or the initial social belief incompatibility related to ϕ, the doctor effectively disregards it, thereby perpetuating the social misalignment from the patient's standpoint.

5 Conclusion

This paper has advanced the understanding of social repair, which is crucial for maintaining interpersonal harmony, by proposing a formal representation centered on the coordination of social beliefs between individuals. Our approach integrates insights from foundational social theories of interaction, principles of social cognition, and established computational frameworks to define core components (e.g., SocialBel, Social interpretation function) and the underlying processes of social repair. We delineated key interactional dynamics where perceived misalignments in social beliefs can trigger observable protests, leading to distinct reactive strategies by interlocutors—namely accommodation, calibration, or ignoring the issue —as illustrated through diverse examples. This framework offers a systematic approach to the social and relational dimensions of repair, providing crucial insights for both dialogue study and the development of more socially intelligent computational agents.

6 Future Work

Building on this paper's insights into strategies for addressing social missteps and maintaining social relationships, future research should extend this foundation in several key directions. These include a deeper investigation into the role of multimodal interactivity in social repair, examining how non-verbal cues such as facial expressions, gestures, and intonation dynamically interact with verbal repair strategies; exploring the longitudinal aspects of social repair to understand the evolution of strategies and their long-term relational impacts; and analyzing cultural and contextual variations to grasp how individuals adapt repair mechanisms across diverse social settings. Furthermore, extending the current dyadic focus to develop models of social repair within more complex multiparty dialogues, a domain where the principles of social repair also apply, is crucial. Finally, developing a more fine-grained taxonomy of social repair types—whether by adapting established distinctions from commu-

nicative repair (such as self-initiated versus other-initiated repair (Schegloff, 1992)) or by establishing new taxonomy based on the nature of social beliefs being repaired or the relational goals being pursued—will significantly enhance both the theoretical understanding and the computational modeling of these nuanced social processes. Incorporating uncertainty is crucial, as strategic ambiguity and plausible deniability are key features of sophisticated social interaction (Pinker et al., 2008), which can be formally achieved by extending our TTR-based model with a probabilistic formulation (Cooper et al., 2015; Noble et al., 2020).

7 Acknowledgement

We thank SemDial anonymous reviewers for their valuable feedback. Alafate Abulimiti was supported by a French government grant managed by the Agence Nationale de la Recherche as part of the France 2030 program, reference ANR-22-EXEN-0004 (PEPR eNSEMBLE / MATCHING). Vladislav Maraev was supported by Swedish Research Council (VR) grant 2023-00358 – Social laughter for virtual agents (SocLaVA).

References

Alafate Abulimiti, Justine Cassell, and Jonathan Ginzburg. 2021. "By the way, do you like Spider Man?" — Towards a social planning model for rapport. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Potsdam, Germany. SEMDIAL.

Ralph Adolphs. 2009. The social brain: neural basis of social knowledge. *Annual review of psychology*, 60(1):693–716.

Penelope Brown and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge university press.

Josep Call and Michael Tomasello. 2008. Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5):187–192.

Herbert H Clark. 1996. *Using language*. Cambridge university press.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication.

Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive science*, 13(2):259–294.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic issues in language technology*, 10.

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

Mark H Davis. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113.

Jean Decety and Philip L Jackson. 2004. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2):71–100.

Mark Dingemanse and N.J. Enfield. 2024. Interactive repair and the foundations of language. *Trends in Cognitive Sciences*, 28(1):30–42.

Kathy Domenici and Stephen W Littlejohn. 2006. *Facework: Bridging theory and practice*. Sage.

Paul Drew. 1997. 'open' class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1):69–101.

Steve Duck. 1994. *Meaningful relationships: Talking, sense, and relating*. Sage Publications, Inc.

Nancy Eisenberg. 2000. Emotion, regulation, and moral development. *Annual review of psychology*, 51(1):665–697.

Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

Joseph P Forgas and Richard Jones. 1985. *Interpersonal behaviour: The psychology of social interaction*. Pergamon Press.

M Georgeff and A Rao. 1991. Modeling rational agents within a bdi-architecture. In *Proc. 2nd Int. Conf. on Knowledge Representation and Reasoning (KR'91). Morgan Kaufmann*, pages 473–484. of.

Jonathan Ginzburg. 2012. *The interactive stance: Meaning for conversation*. Oxford University Press.

Jonathan Ginzburg, Ivan Sag, and Matthew Purver. 2003. Integrating conversational move types in the grammar of conversation. *Perspectives on dialogue in the new millennium*, 114:25–42.

Erving Goffman. 1959. The presentation of self in everyday life. In *Social theory re-wired*, pages 450–459. Routledge.

Erving Goffman. 1967. *Interaction ritual: Essays in face-to-face behavior*. Routledge.

Stephanie Gross, Brigitte Krenn, and Matthias Scheutz. 2017. The reliability of non-verbal cues for situated reference resolution and their interplay with language: implications for human robot interaction. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*.

Michael Guilfoyle. 2003. Dialogue and power: A critical analysis of power in dialogical therapy. *Family process*, 42(3):331–343.

Antonia F de C Hamilton and Judith Holler. 2023. Face2face: advancing the science of social interaction.

Peter A Heeman and James Allen. 1999. Speech repains, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue. *Computational Linguistics*, 25(4):527–572.

Martin L Hoffman. 1996. Empathy and moral development. *The annual report of educational psychology in Japan*, 35:157–162.

Paul Hömke, Stephen C Levinson, Alexandra K Emmendorfer, and Judith Holler. 2025. Eyebrow movements as signals of communicative problems in human face-to-face interaction. *Royal Society Open Science*, 12(3):241632.

Jiefeng Jiang, Bohan Dai, Danling Peng, and Qian Cui. 2021. A hierarchical model for interpersonal verbal communication. *Social Cognitive and Affective Neuroscience*, 16(1-2):246–256.

Willem JM Levelt. 1983. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104.

Bill Noble, Ellen Breitholtz, and Robin Cooper. 2020. Personae under uncertainty: the case of topoi. In *Proceedings of the Probability and Meaning Conference (PaM 2020)*, pages 8–16.

John Oetzel, Stella Ting-Toomey, Tomoko Masumoto, Yumiko Yokochi, Xiaohui Pan, Jiro Takai, and Richard Wilcox. 2001. Face and facework in conflict: A cross-cultural comparison of china, germany, japan, and the united states. *Communication monographs*, 68(3):235–258.

Alex Sandy Pentland. 2012. The new science of building great teams. *Harvard business review*, 90(4):60–69.

Steven Pinker, Martin A Nowak, and James J Lee. 2008. The logic of indirect speech. *Proceedings of the National Academy of sciences*, 105(3):833–838.

Matthew Purver, Julian Hough, and Christine Howes. 2018. Computational models of miscommunication phenomena. *Topics in cognitive science*, 10(2):425–451.

Elizabeth Redcay, David Dodell-Feder, Mark J Pearrow, Penelope L Mavros, Mario Kleiner, John DE Gabrieli, and Rebecca Saxe. 2010. Live face-to-face interaction during fmri: a new tool for social cognitive neuroscience. *Neuroimage*, 50(4):1639–1647.

Emanuel A Schegloff. 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American journal of sociology*, 97(5):1295–1345.

Emanuel A Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Helen Spencer-Oatey. 2005. (im) politeness, face and perceptions of rapport: unpackaging their bases and interrelationships.

Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293.

David Traum. 1994. A computational theory of grounding in natural language conversation.

David Traum. 2003. Issues in multiparty dialogues. In *Workshop on Agent Communication Languages*, pages 201–211. Springer.

Frank Van Overwalle. 2009. Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3):829–858.

Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759.

Henry M Wellman. 2018. Theory of mind: The state of the art. *European Journal of Developmental Psychology*, 15(6):728–755.

Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents: 14th International Conference, IVA 2014, Boston, MA, USA, August 27-29, 2014. Proceedings 14*, pages 514–527. Springer.