Transactions on Social Science, Education and Humanities Research
ISSN: 2960-1770, eISSN: 2960-2262 | Volume 16
EPHHR 2026

From "Talking at Cross Purposes" to "Tacit Understanding": The Psycholinguistic Mechanisms of Pragmatic Inference

Shuning Wang

School of Foreign Languages, Shandong Normal University, Jinan, China

wsnwangshuning@163.com

Abstract. This study develops an integrative framework for pragmatic inference—the capacity to infer speaker intentions beyond literal meaning. Integrating the Cooperative Principle, Relevance Theory, Theory of Mind, and executive functions, it conceptualizes inference as a dynamic, resource-dependent process. Success depends on contextual alignment, accurate intention attribution, and cognitive resources. Incorporating repair mechanisms from Conversation Analysis, the framework reveals how interactants achieve understanding through iterative strategies such as reformulation and clarification. The paper elucidates the pathway from "talking at cross purposes" to "tacit understanding," providing theoretical synthesis and practical guidance for communicative success.

Keywords: pragmatic inference; Relevance Theory; Theory of Mind; executive functions; repair mechanism.

1. Introduction

From "talking at cross purposes" to "tacit understanding"—the distance between the two is spanned by the most central cognitive capacity in human linguistic communication: pragmatic inference. Pragmatic inference refers to the cognitive process whereby the hearer goes beyond literal meaning, drawing on the immediate context and background knowledge to infer the speaker's real intention. Grice (1957) distinguished "natural meaning" from "non-natural meaning"[1], pointing out that the essence of verbal communication lies in intention recognition: the speaker displays a communicative intention to the hearer, and the hearer, in turn, infers the "implicature" by relying on the shared, default Cooperative Principle and its conversational maxims. Grice (1975) systematically elaborated this idea[2], laying the theoretical cornerstone of modern pragmatics and establishing inference as the core concept for understanding linguistic communication.

Grice's theory, however, does not fully explain the cognitive mechanisms of intention recognition. Sperber and Wilson's (1986)[3] Relevance Theory advanced this line of inquiry by grounding pragmatic inference in human cognitive architecture: cognition inherently pursues maximal relevance, while ostensive-inferential communication automatically conveys a presumption of optimal relevance—that the utterance yields adequate contextual effects and is worth the hearer's processing effort. The hearer tests possible interpretations in order of decreasing relevance; once an interpretation meets the criterion of sufficient relevance, inference stops. In this way, pragmatic inference was shifted from a socio-normative plane to a cognitive-psychological one, becoming a basic property of human cognition.

Unlike theories centred on successful inference, Healey, de Ruiter, and Mills (2018)[4]adopt a perspective of "failure". They observe that traditional research treats successful communication as the default and regards misunderstanding merely as noise external to the system, whereas it is precisely the analysis of misunderstanding and repair processes that can reveal the dynamic nature of pragmatic inference in real interaction. This shift in perspective has opened new directions for research.

The rise of cognitive approaches has fueled a flourishing of empirical studies. A longitudinal study by Gabbatore et al. (2025)[5] found that 3- to 8-year-old children's ability to understand complex communicative intentions such as irony and deception is closely related to executive functions (e.g.,

inhibitory control, working memory) and Theory of Mind development, and that these associations display differentiated patterns depending on the type of intention—indicating that pragmatic inference is a complex enterprise involving the coordination of multiple cognitive systems. Bohn and Frank (2025)[6] experimentally demonstrated that adults' pragmatic inference from nonverbal behaviours such as gaze and gesture likewise rests on sophisticated inferences about the speaker's knowledge state and intentions. Approaching the issue from the autism spectrum, Dindar et al. (2022)[7] found that difficulties in pragmatic inference are linked to visual attention timescales and physiological arousal levels, further corroborating the idea of multi-system coordination. Together, these studies show that pragmatic inference is a psychological reality that can be precisely measured and experimentally investigated, and that its realisation depends on the integration of cognitive, attentional, and emotional resources.

To compensate for the limited ecological validity of laboratory studies, scholars have turned their attention to broader real-world communicative settings. Jiang (2017)[8]emphasised that inference models based purely on informational quantity, divorced from specific sociocultural contexts, cannot genuinely account for the complexity and flexibility of pragmatic inference, and proposed a "socio-cognitive" model that integrates situational factors such as intention, knowledge, emotion, and identity. This perspective resonates with the tradition of Interactional Linguistics: Elder and Haugh (2024)[9] noted that pragmatic inference involves not only reconstructing the speaker's intention but also the avoidance and negotiation of "unwanted inferences" in interaction. Through micro-analysis of live-streaming interaction, Song and Licoppe (2024)[10] revealed "attention-based behaviour" as a core resource for understanding the streamer's intentions, extending the research to multimodal, asymmetric communicative settings. Kelly and Miller-Ott (2018)[11], analysing mobile phone text messages, found that the absence of non-verbal cues easily triggers referential and affective misunderstandings, providing empirical evidence for the influence of the media environment on pragmatic inference. These analyses offer a window for observing the dynamic workings of the inference process.

At the same time, research has increasingly attended to the interplay between pragmatic inference and other levels of language processing as well as to cross-linguistic differences. Ronai et al. (2019)[12], using eye-tracking, showed that pragmatic inference exerts only a limited influence on speech perception, and that this influence occurs at a late stage of processing, revealing the dynamic integration of information across different levels. In a cross-linguistic comparison, Carneiro et al. (2020)[13] found that levels of false memory for pragmatically implicative sentences vary significantly according to language and cultural background, urging caution in the cross-cultural application of pragmatic inference materials. Filipovic (2022)[14], comparing British and American police interrogations, demonstrated that in high-stakes, power-asymmetrical institutional dialogues, the success or failure of pragmatic inference directly bears on the validity of legal evidence and even on judicial fairness, highlighting its real-world significance.

In summary, research on pragmatic inference has progressed from philosophical concept analysis to the exploration of cognitive-psychological mechanisms, then extended to the dynamic analysis of social interaction, and further expanded into cross-cultural, multimodal, and institutional dialogue settings. These strands of research, however, have largely proceeded in isolation, lacking a unified framework that integrates theoretical insights, empirical data, and interactional observations. How pragmatic inference is implemented in the brain, how core cognitive resources such as executive functions, Theory of Mind, and social-context processing are dynamically recruited, and where the explanatory boundaries of different theoretical models lie—all these questions await systematic clarification.

It is against this backdrop that the present study focuses on the psycholinguistic mechanisms of pragmatic inference, exploring how human beings move beyond literal meaning to achieve "tacit understanding". The paper first reviews core theoretical models, and then distils key psychological mechanisms—such as Theory of Mind, executive functions, and the moderating role of social-contextual factors—from cognitive and developmental psychology research. Through a contrastive

analysis of the two polar phenomena of “talking at cross purposes” and “tacit understanding”, this study aims to reveal the precise and efficient cognitive landscape behind linguistic communication, to provide a scholarly reference for the psychological reality of pragmatic inference, and to attempt to construct a more explanatory integrative framework.

2. Theoretical Framework

The cognitive mechanisms of pragmatic inference require explanation from multiple theoretical dimensions. This study selects the Cooperative Principle, Relevance Theory, and Theory of Mind and executive functions as its core analytical tools, corresponding respectively to the socio-normative foundation, the cognitive optimality mechanism, and the mental resource guarantee along the chain of pragmatic inference. This section will sequentially expound the core claims of the three theories and their explanatory power for pragmatic inference, then illustrate the organic interconnections among them, and, on this basis, construct a mechanism analysis framework for the success and failure of pragmatic inference. The overall logical structure of this framework is shown in Figure 1.

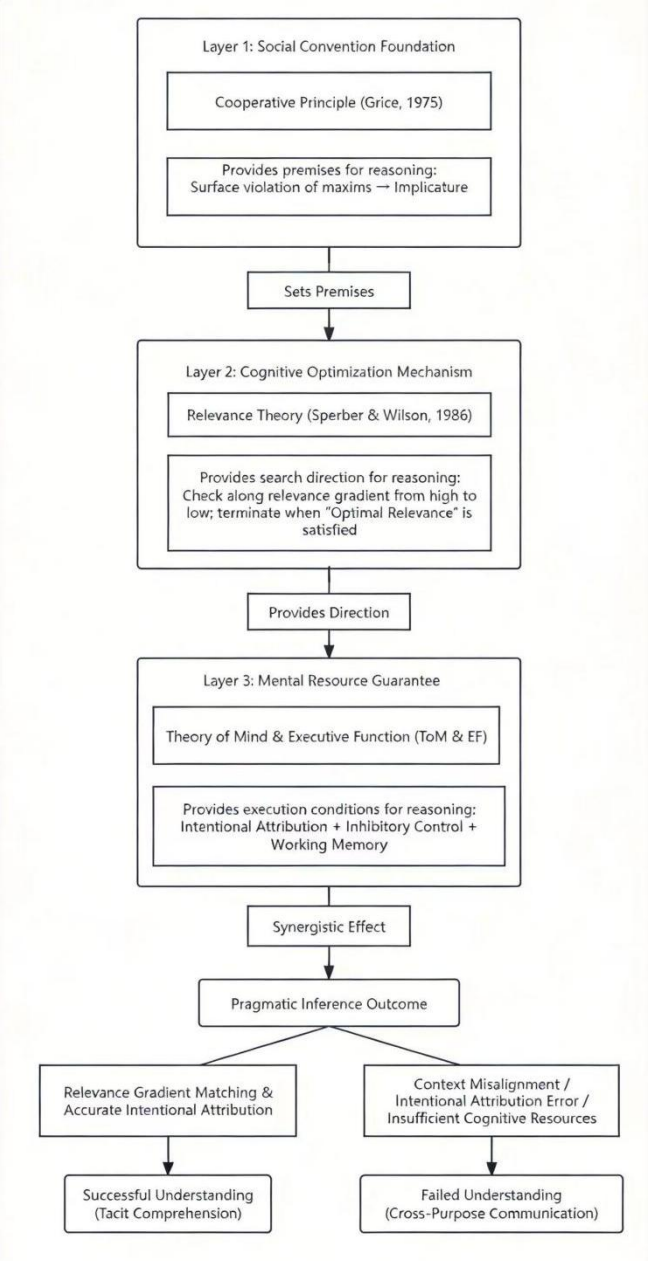

Figure 1. An integrative analysis framework of the mechanisms of success and failure in pragmatic inference

2.1. The Cooperative Principle: The Socio-normative Basis of Inference

The Cooperative Principle provides the most fundamental presupposition for pragmatic inference. Grice (1975) pointed out that participants in verbal communication, by default, jointly observe a set of conversational norms, including the maxim of quantity (provide the required amount of information, and no more), the maxim of quality (speak truthfully and with adequate evidence), the maxim of relation (say things that are relevant to the topic), and the maxim of manner (express oneself clearly and in an orderly manner). The core function of this principle is as follows: when a speaker ostensibly violates a maxim, the hearer does not consider the speaker uncooperative but instead infers a specific communicative intention behind it, thereby deriving a "conversational implicature." For instance, when asked for an evaluation of a new proposal, if the reply is "I have a bit of a cold," it ostensibly violates the maxim of relation, yet the hearer can infer that the speaker intends to tactfully avoid offering an evaluation. The Cooperative Principle reveals the socio-contractual nature that makes pragmatic inference possible: inference is possible because both parties share a set of default expectations about "how to speak." However, this principle mainly answers the question of "on what basis does inference rest," without explaining how the hearer rapidly pinpoints the correct one among multiple possible implicatures—a crucial gap that subsequent theories need to fill.

2.2. Relevance Theory: The Cognitive Optimality Mechanism of Inference

To address the limitations of the Cooperative Principle in explaining the inferential process, Relevance Theory offers a more refined mechanistic account from a cognitive-psychological perspective. Sperber and Wilson (1986) argued that the human cognitive system operates with "maximal relevance" as its goal, i.e., to achieve the greatest contextual effects for the least cognitive effort. In verbal communication, an utterance automatically conveys a presumption of "optimal relevance": the utterance yields adequate contextual effects and is worth the hearer's processing effort. The hearer tests possible interpretations in order of decreasing relevance; once an interpretation meets the criterion of "sufficient relevance," inference stops there, and other possibilities are not considered. Taking "It's a bit stuffy in here" as an example, the hearer will prioritize testing the most relevant interpretation among possibilities such as "poor air circulation," "feeling uncomfortable," and "wanting to open the window"—the last one is selected because it can guide subsequent action and thus yields stronger contextual effects. This mechanism explains how pragmatic inference can be completed in a very short time: it is not exhaustive deductive reasoning over all possibilities but a rapid lock onto the optimal solution under the constraints of cognitive resources. Yet Relevance Theory likewise presupposes that the hearer possesses normal mental state attribution abilities and sufficient cognitive resources—precisely the factors that sometimes cause pragmatic inference to fail.

2.3. Theory of Mind and Executive Functions: The Subject Conditions for Inference

The Cooperative Principle and Relevance Theory answer the questions of "on what basis does inference rest" and "how does inference select the optimal interpretation," yet they presuppose a cognitive subject with fully intact mental capacities. In reality, the realization of pragmatic inference depends on specific psychological abilities. Theory of Mind refers to the ability to infer one's own and others' mental states (beliefs, intentions, emotions, etc.)—this is the core of intention recognition. Empirical research by Gabbatore et al. (2025) demonstrated that children's ability to understand complex pragmatic phenomena such as irony and deception is significantly correlated with the developmental level of Theory of Mind, and that the degree of reliance on Theory of Mind varies across different types of pragmatic tasks. Furthermore, inhibitory control and working memory, two core components of executive functions, play a moderating role in the inferential process: inhibitory control helps the hearer eliminate irrelevant candidates among multiple possible interpretations, while working memory sustains the online processing of contextual information. Bohn and Frank's (2025) experiment further demonstrated that even pragmatic inference from nonverbal communicative behaviours (such as gaze and gesture) relies on sophisticated inferences about the speaker's knowledge state and behavioural intentions. It can thus be seen that Theory of Mind and executive

functions constitute the "hardware conditions" for successful pragmatic inference—when these abilities are insufficient, inference may still fail even if the Cooperative Principle and the relevance mechanism function normally.

2.4. The Organic Interconnection of the Three Theories: From Separation to Synergy

The three theoretical tools described above are neither independent of nor substitutes for one another; rather, each corresponds to a different link in the chain of pragmatic inference, and together they form the basic explanatory loop for pragmatic inference. The Cooperative Principle provides the normative premise for inference: both parties assume by default that the speaker is cooperative, so that apparent deviations in informativeness, clarity, or relevance of an utterance should often be understood as strategic expressions of intention rather than meaningless aberrations. Building on this, Relevance Theory further reveals the optimality mechanism of inference: the hearer tends to test candidate interpretations in order of decreasing relevance and terminates the search once "sufficient relevance" is achieved, thereby completing meaning derivation within a limited time. Meanwhile, research on Theory of Mind and executive functions supplements the account from the angle of subject abilities, specifying the resource conditions for smooth inference—namely, the hearer needs the capacity to attribute beliefs and intentions to others and depends on inhibitory control and working memory to maintain contextual information and exclude irrelevant interpretations. A deficiency in any link can lead to misunderstanding: for instance, a breakdown of the cooperative expectation weakens the basis for inference, inappropriate assessment of relevance can skew the direction of interpretation, and limited Theory of Mind or executive functions may prevent the inferential process from being completed.

It should be pointed out, however, that these three theories have different explanatory emphases, and their explanatory power is bounded. The Cooperative Principle mainly remains at the normative level, offering little detail on the processing specifics of how candidate interpretations are generated and filtered. Relevance Theory, although emphasizing cognitive economy and specifying an inferential search path, still falls short in accounting for social-structural factors in power-asymmetrical or institutionalized interactions. Research on Theory of Mind and executive functions can reveal sources of individual differences but rarely addresses how, once misunderstanding occurs, interacting parties achieve meaning repair through clarification, reformulation, and negotiation. Therefore, to capture more fully the mechanisms of success and failure in pragmatic inference in real-life communication, it is necessary to achieve further integration at the three levels of normative presupposition, cognitive optimality, and interactional repair.

2.5. Mechanism Analysis: A Classification Framework for Success and Failure in Pragmatic Inference

Based on the above theoretical integration, the cognitive mechanism of pragmatic inference can be summarized as the following closed-loop process: The speaker produces an utterance → This utterance carries varying degrees of relevance depending on the context and the manner of expression → The hearer, constrained by his or her own cognitive resources, performs attributional inference over possible intentions in order of decreasing relevance → If a given interpretation meets the criterion of "sufficient relevance," inference succeeds and "tacit understanding" is achieved; if relevance assessment fails or intention attribution is erroneous, misunderstanding occurs, i.e., "talking at cross purposes." Accordingly, comprehension failure can be attributed to three levels: contextual misalignment (inappropriate relevance assessment), erroneous intention attribution (limitations in Theory of Mind ability), and insufficient cognitive resources (executive function constraints). As shown in Table 1, the success and failure of pragmatic inference can be classified along three dimensions: the performance level, the mechanism level, and the cause level.

Table 1. A three-level classification framework for success and failure in pragmatic inference

Outcome Level	Mechanism Level	Cause Level
Successful understanding (tacit understanding)	Relevance gradient matching + accurate intention attribution	Speaker's expression is sufficient, context is accessible; listener's theory of mind and executive functions are normal
Understanding failure: context misalignment	Improper relevance assessment	Speaker's unclear communication (unclear knowledge or vague expression); listener's insufficient contextual knowledge base
Understanding failure: intention attribution error	Mental state attribution bias	Listener's insufficient theory of mind ability, unable to correctly infer the speaker's beliefs or intentions
Understanding failure: insufficient cognitive resources	Inhibition failure or working memory overload	Listener's executive functions are limited, unable to maintain contextual information or filter out irrelevant interpretations

The following everyday dialogue example illustrates the above mechanism. Suppose speaker A says to B in a sealed room, "It's a bit stuffy in here." From the perspective of the Cooperative Principle, A's utterance ostensibly states an objective fact; but placed in the context of "the room is sealed," hearer B will assume by default that A is cooperative, so the utterance should carry a communicative intention beyond the literal meaning. From the perspective of Relevance Theory, B will prioritize testing the most relevant interpretation: among the possibilities of "poor air circulation," "feeling uncomfortable," and "wanting to open the window," the last one is selected as the most relevant because it can guide subsequent action and thus yields stronger contextual effects. From the perspective of Theory of Mind and executive functions, B needs to infer that A's intention is not to state a fact but to issue an indirect request (Theory of Mind), while also inhibiting the irrelevant interpretation of "just making a casual remark" and maintaining the contextual information "the room is sealed" (executive function). If all these links function normally, B successfully infers the request "open the window" and achieves "tacit understanding." Conversely, if B lacks the relevant contextual knowledge (e.g., does not know the room is sealed), or has insufficient Theory of Mind ability (unable to recognize the intention of an indirect request), or suffers from working memory overload (forgets the previous description of the room), then B may remain at the literal level of understanding, resulting in "talking at cross purposes." This example demonstrates that the Cooperative Principle, Relevance Theory, and Theory of Mind/executive functions do not operate in isolation in pragmatic inference but work in synergy to form a complete explanatory chain that runs from social convention, through cognitive optimality, to mental resource guarantees.

2.6. From One-Shot Inference to Dynamic Repair: The Revisability of Pragmatic Inference

The above mechanism analysis implies an important assumption: pragmatic inference is a one-shot process—upon receiving an utterance, the hearer tests interpretations along a gradient of relevance, and it either succeeds or fails. Real verbal communication, however, is not like this. When initial inference fails, communication does not simply terminate. Instead, both parties, through a series of interactive behaviours, readjust the context set and the ordering of relevance, initiating "second-round inference," thereby achieving repair of comprehension failure.

This shift in perspective owes much to long-standing research on the "repair" mechanism in Conversation Analysis. Schegloff, Jefferson, and Sacks (1977) [15] were the first to systematically describe the structural procedures through which people deal with hearing and understanding

problems in conversation, pointing out that repair is not an anomalous event but a routine component of conversational organization. Healey, de Ruiter, and Mills (2018) further pointed out that it is precisely the analysis of misunderstanding and repair processes that can reveal the dynamic nature of pragmatic inference in real interaction. In other words, pragmatic inference should not be understood as a single-shot, one-way decoding of intentions but as a dynamic process that can be iterated and negotiated.

Based on this understanding, this section will elaborate, from the perspectives of the speaker and the hearer, on the specific behaviours each can adopt to "restart" the inferential process when initial inference fails.

2.6.1. The Speaker's Adjustment Strategies: Reformulation, Context Specification, and Relevance Modulation

When a speaker realizes that the hearer has failed to correctly understand the intended meaning (usually detected through the other's response—questions, silence, off-topic answers, etc.), the speaker can actively adjust his or her utterance output to reduce the hearer's cognitive load or correct the hearer's relevance assessment. Specific strategies include:

First, reformulation. The speaker re-expresses the same intention in a more explicit and direct manner. For example, after "It's a bit stuffy in here" is misunderstood as merely stating the air condition, the speaker can add: "Open the window." In this case, the originally implicit request intention is made explicit, and the hearer no longer needs to perform indirect inference.

Second, context specification. The speaker supplements contextual information that the hearer may be lacking. For example, if the hearer does not know the room is sealed, the speaker can say: "This room has no windows, there's no ventilation." Such supplementation directly acts on the hearer's contextual knowledge base, thereby correcting the direction of relevance assessment.

Third, relevance modulation. The speaker re-calibrates the most relevant components of the utterance through pragmatic markers (such as "actually," "I mean") or by changing tone and pace. Relevance Theory notes that relevance is a matter of degree; the speaker can use linguistic means to boost the relevance gradient of a specific interpretation, so that it is tested earlier in the hearer's search path.

2.6.2. The Hearer's Proactive Strategies: Clarification Requests, Hypothesis Testing, and Context Expansion

When inference is blocked, the hearer does not passively wait but can actively intervene in the repair process through the following behaviours:

First, issuing a clarification request. This is the most direct way to initiate repair. The hearer can ask "What do you mean?", "Are you saying…?", etc., requesting more information from the speaker. Research by Purver, Ginzburg, and Healey (2003)[16] showed that clarification requests occur frequently in natural conversation, and that their form (e.g., specific question words, partial repeats) shows systematic correspondences with types of misunderstanding.

Second, hypothesis testing and confirmation. The hearer formulates a conjectured interpretation in a tentative manner and asks the speaker to confirm or deny it. For example: "Do you mean I should open the window?" Here, the hearer actively offers a candidate interpretation, and the speaker makes a correctness judgment—this essentially shifts part of the inferential burden to the speaker, reaching consensus through interactive negotiation.

Third, proactive context expansion. The hearer actively retrieves or asks about potentially missing contextual information. For example: "Does this room have no windows?" Such behaviour directly targets the contextual knowledge base, correcting the premises for inference.

The effectiveness of these strategies has been confirmed by multiple empirical studies. Elder and Haugh (2024), in their editorial introduction, pointed out that pragmatic inference is not only about how the hearer reconstructs the speaker's inner intention but also manifests in the avoidance and

negotiation of "unwanted inferences" during interaction. Song and Licoppe (2024), through micro-analysis of live-streaming interaction, showed how viewers use "attention-based behaviour" to actively adjust their inferential paths regarding the streamer's intentions. Kelly and Miller-Ott (2018) found in their study of mobile phone text messages that, when a lack of nonverbal cues leads to comprehension failure, both parties often repair misunderstandings by appending explanations or repeating formulations. Together, these studies indicate that the repair mechanism of pragmatic inference displays cross-contextual universality across different media and settings.

2.7. An Integrative Model of the Repair Mechanism: The Second-Round Inference Loop

Based on the above analysis, the repair mechanism of pragmatic inference can be integrated into an expanded closed-loop model. Building on the original "one-shot inference," this model adds an iterative repair loop of "failure detection—repair initiation—second-round inference," as illustrated in Figure 2.

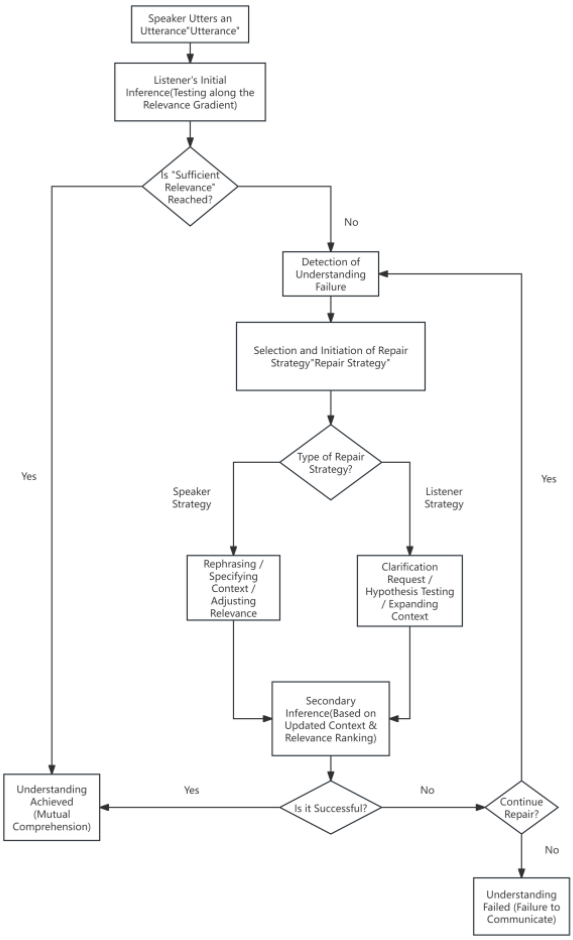

Figure 2. The iterative closed-loop model of pragmatic inference repair

Initial stage: The speaker produces an utterance, and the hearer performs first-round inference along the relevance gradient. If the correct interpretation is successfully matched at the point of "sufficient relevance," "tacit understanding" is achieved, and the inferential process terminates.

Failure detection: If the first-round inference fails to produce an interpretation consistent with the speaker's intention (manifested as a hearer response that does not match the speaker's expectation, or the hearer actively expresses non-understanding), then both parties or one of them become aware of a possible comprehension failure.

Repair initiation: Depending on the specific situation and roles, the speaker or the hearer (or both) selects an appropriate repair strategy. The speaker can adopt reformulation, context specification, or relevance modulation; the hearer can issue a clarification request, engage in hypothesis testing, or

proactively expand the context. The choice of repair strategy is influenced by multiple factors, including the relationship between the parties (e.g., power asymmetry), medium characteristics (e.g., face-to-face vs. text messaging), and the cognitive resources of both sides.

Second-round inference: On the basis of the new information provided by the repair behaviour (such as more explicit expression, supplementary context, or confirmed interpretation), the hearer re-inferences along the relevance gradient. At this point, since the repair behaviour has substantially boosted the relevance of the target interpretation (e.g., reformulation makes the implicit intention explicit), the success rate of the second-round inference is significantly higher than that of the first.

Continued iteration: If the second-round inference still fails, both parties can continue the above loop until understanding is achieved or one side gives up (see Table 2 for the classification of repair strategies). Filipovic's (2022) study of British and American police interrogations showed that in institutional dialogues, whether the repair loop persists is often constrained by institutional goals—for example, in adversarial American interrogations, the police sometimes deliberately refrain from initiating repair in order to maintain an ambiguous interpretation that is unfavourable to the suspect.

Table 2. Classification of repair strategies for pragmatic inference

Repair Initiator	Strategy Type	Specific Behavior	Mechanism
Speaker	Reformulation	Re-express same intention more explicitly	Reduce inferential depth; make implicit intention explicit
Speaker	Context Elaboration	Supply missing contextual information	Expand listener's context knowledge; correct relevance assessment
Speaker	Relevance Adjustment	Raise relevance gradient of target interpretation via pragmatic markers/tone	Alter relevance ranking; test target interpretation earlier
Listener	Clarification Request	Ask "What do you mean?" or similar	Request more information; trigger speaker repair
Listener	Hypothesis Testing	Express tentative interpretation and request confirmation	Shift inference burden to speaker; negotiate consensus
Listener	Context Expansion	Actively retrieve or inquire about missing contextual information	Correct inference premises; fill knowledge gaps

2.8. The Supplementary Significance of the Repair Mechanism

Incorporating the repair mechanism into the theoretical framework carries threefold supplementary significance. First, it revises the "one-shot" assumption of pragmatic inference and better aligns it with the dynamic character of real communication. In real interaction, understanding is not achieved all at once but is gradually attained through repeated trial and adjustment. Second, it clarifies the distinct roles and responsibilities of the speaker and the hearer in the repair process. Traditional pragmatic inference theories have often emphasized the hearer's inferential activity, but the analysis of repair mechanisms shows that the speaker also has a responsibility to actively adjust expressions to lower the comprehension threshold—a point particularly salient in Filipovic's (2022) comparative study of British and American police interrogations: British police officers were more inclined to actively reformulate or specify questions when understanding was blocked, whereas American officers were more likely to maintain vague formulations to secure interpretations unfavourable to the suspect. Third, it offers operationalizable behavioural pathways for "remedying" pragmatic inference failure. The six repair strategies summarized in Table 2 can be translated into concrete communication advice: when a speaker realizes the other party may not understand, they should promptly reformulate or supplement the context; when a hearer is confused, they should actively issue clarification requests or engage in hypothesis testing. The effectiveness of these strategies has been corroborated across diverse communicative settings (everyday conversation, text messaging, live-streaming interaction, institutional interviews).

In summary, pragmatic inference is not a one-shot, one-way cognitive decoding but a dynamic, negotiative process that is feedback-friendly, iterable, and repairable. The Cooperative Principle sets the premise of inference, Relevance Theory guides its direction, Theory of Mind and executive functions supply the resources, and the repair mechanism provides a "second chance" when inference fails. The integration of these four components constitutes the full explanatory pathway from "talking at cross purposes" to "tacit understanding." The concluding section that follows will further summarize the present study's contribution to the theoretical integration of pragmatic inference and offer research prospects and practical recommendations for the future.

3. Conclusion

This study set out to address a core question: How do humans go beyond literal meaning in verbal communication and infer the speaker's real intention? Approaching this question from a psycholinguistic perspective, the paper constructed an integrative theoretical analysis framework and, on that basis, systematically explored the mechanisms of success and failure in pragmatic inference.

First, this study proposed a workable scheme for theoretical integration. Existing research presents a divided landscape at the theoretical level: the Cooperative Principle reveals the socio-normative premise of inference, Relevance Theory illuminates the cognitive optimality pathway, and research on Theory of Mind and executive functions uncovers the mental resource conditions. Each of the three explains a separate link in the chain of pragmatic inference, yet they lack organic integration. Taking "socio-normative foundation – cognitive optimality mechanism – mental resource guarantee" as its logical thread, this paper incorporated the above theories into a unified framework and further introduced the repair mechanism from Conversation Analysis, pointing out that pragmatic inference is not a one-shot decoding process but a dynamic, negotiative process that is feedback-friendly, iterable, and repairable.

Second, this paper conducted a mechanism-level classification analysis of success and failure in pragmatic inference. Building on the theoretical integration, the cognitive process of inference is summarized as follows: the speaker produces an utterance → the utterance carries different relevance gradients depending on context and manner of expression → the hearer, constrained by cognitive resources, tests possible interpretations of intention along a gradient from high to low relevance → if an interpretation meets the criterion of "sufficient relevance," inference succeeds; otherwise, comprehension deviation occurs. Accordingly, comprehension failure is attributed to three levels: contextual misalignment (inappropriate relevance assessment), erroneous intention attribution (limitations in Theory of Mind ability), and insufficient cognitive resources (executive function constraints), and these were systematically classified along the performance, mechanism, and cause levels by means of a three-line table, providing a theoretical basis for defining comprehension failure in subsequent empirical research.

Third, this paper offered actionable recommendations for the repair process following inference failure. In real communication, the failure of initial inference does not mean the end of communication. From the perspectives of the speaker and the hearer, six types of repair strategies were identified: the speaker can reformulate, specify the context, or modulate relevance; the hearer can issue clarification requests, conduct hypothesis testing, or proactively expand the context. These strategies differ in their operating mechanisms and function respectively in making intentions explicit, expanding the knowledge base, altering the search ordering, triggering repair, shifting the inferential burden, and correcting preconditions. Furthermore, this paper constructed an iterative closed-loop model of "failure detection – repair initiation – second-round inference," pointing out that the updated information provided by repair behaviour enables the hearer to re-infer, and the success rate of second-round inference is significantly higher than that of the first round.

Finally, this study offered prospects for theoretical development and practical application. At the theoretical level, the integrative framework still requires empirical testing: whether different types of comprehension failure differ in their neural representations, and how the cognitive costs and

efficiency of different repair strategies can be quantitatively compared—these questions await answers from behavioural experiments, eye-tracking, and neuroimaging studies. At the practical level, the classification of repair strategies can be directly translated into communication training content, enabling the design of training programmes for police interrogations, doctor-patient communication, psychological counselling, and other high-stakes communication professions, thereby reducing communication breakdowns caused by inference failure; it can also be used to design stepwise training for children with delayed pragmatic development or individuals on the autism spectrum, gradually enhancing their inferential ability.

In summary, pragmatic inference is not a one-shot decoding confined within the individual mind, but a dynamic, negotiative process embedded in interaction, reliant on cognitive resources, and made iterable through the repair mechanism. By integrating the Cooperative Principle, Relevance Theory, Theory of Mind and executive functions, and the repair mechanism, this study has constructed the full explanatory pathway from "talking at cross purposes" to "tacit understanding," hoping to provide an analytical tool for subsequent empirical research and a scholarly reference for coping with comprehension failures in real communication.

References

[1] Grice, H. P. (1957). Meaning. The Philosophical Review, 66(3), 377–388. https://doi.org/10.2307/2182440

[2] Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics, Vol. 3: Speech acts (pp. 41–58). Academic Press.

[3] Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Blackwell.

[4] Healey, P. G. T., de Ruiter, J. P., & Mills, G. J. (2018). Editors' introduction: Miscommunication. Topics in Cognitive Science, 10(2), 264–278. https://doi.org/10.1111/tops.12340

[5] Gabbatore, I., Hilviu, D., Merlini, A., & Bosco, F. M. (2025). Pragmatic comprehension and inferential ability: Interplay with executive functions and Theory of Mind. Philosophical Transactions of the Royal Society B: Biological Sciences, 380(1932), 20230491. https://doi.org/10.1098/rstb.2023.0491

[6] Bohn, M., & Frank, M. C. (2025). Pragmatics as social inference about intentional action. Open Mind, 9, 290–304. https://doi.org/10.1162/opmi_a_00191

[7] Dindar, K., Loukusa, S., Helminen, T. M., Mäkinen, L., Siipo, A., Laukka, S., Rantanen, A., Mattila, M. L., Hurtig, T., & Ebeling, H. (2022). Social-pragmatic inferencing, visual social attention and physiological reactivity to complex social scenes in autistic young adults. Journal of Autism and Developmental Disorders, 52(7), 73–88. https://doi.org/10.1007/s10803-021-05182-3

[8] Jiang, W. (2017). A socio-cognitive approach to pragmatic inference. Intercultural Pragmatics, 14(3), 421. https://doi.org/10.1515/ip-2017-0017

[9] Elder, C.-H., & Haugh, M. (2024). The role of inference and inferencing in pragmatic models of communication. Journal of Pragmatics, 229, 71–76. https://doi.org/10.1016/j.pragma.2024.06.002

[10] Song, L., & Licoppe, C. (2024). Noticing-based actions and the pragmatics of attention in expository live streams. Noticing 'effervescence' and noticing-based sequences. Journal of Pragmatics, 226, 1–16. https://doi.org/10.1016/j.pragma.2024.04.001

[11] Kelly, L., & Miller-Ott, A. E. (2018). Perceived miscommunication in friends' and romantic partners' texted conversations. Southern Communication Journal, 83(4), 267–280. https://doi.org/10.1080/1041794X.2018.1488271

[12] Ronai, E., Sun, Y., Yu, A. C. L., & Xiang, M. (2019). Integration of contextual-pragmatic and phonetic information in speech perception: An eye-tracking study. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10(1), 1–15. https://doi.org/10.5334/labphon.186

[13] Carneiro, P., Lapa, A., Reis, J., & Ramos, T. (2020). Testing pragmatic inferences: The impact of language and culture. Psicológica, 41(1), 1–20. https://doi.org/10.2478/psicolj-2020-0001

[14] Filipović, L. (2022). The good, the bad and the ugly: Miscommunication in UK police interviews and US police interrogations. Journal of Police and Criminal Psychology, 37(2), 297–311. https://doi.org/10.1007/s11896-022-09495-w

[15] Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. Language, 53(2), 361–382. https://doi.org/10.2307/413107

[16] Purver, M., Ginzburg, J., & Healey, P. G. T. (2003). On the means for clarification in dialogue. In R. Smith & J. van Kuppevelt (Eds.), Current and new directions in discourse and dialogue (pp. 235–256). Kluwer Academic Publishers.