ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

A Theoretical Note on the Cognitive Payoff of Indirect Speech

Zhou Jianhua[1,a,*]

[1]*Lyceum of the Philippines University, Manila, 1002, Philippines*
[a]*191517392@qq.com*
[*]*Corresponding author*

Abstract: *According to Relevance Theory, indirect speech can produce richer cognitive effects to offset the additional processing effort. However, most empirical studies have been limited to propositional reasoning, while the exploration of the cognitive effects of other forms of indirect speech is insufficient. Through the conceptual integration of experimental results and discourse data, this note proposes a four-level typology of additional cognitive effects: interpersonal, humorous, aesthetic, and informational. By re-incorporating Brown and Levinson's politeness theory and Attardo's humor framework into the Relevance Theory model, it is shown that apparent irrelevance can often be compensated for by multiple social and emotional benefits. Finally, this note puts forward two testable hypotheses for future experimental research, emphasizing that even without increasing propositional content, optimal relevance can be achieved through emotional and social benefits.*

Keywords: *Relevance Theory; Indirect speech; Cognitive effects; Politeness; Humour; Aesthetics*

1. Introduction

According to Relevance Theory (RT), every utterance triggers an expectation for the optimal degree of relevance: the hearer is entitled to achieve sufficient cognitive effect without incurring unnecessary cognitive effort [1]. For instance, when the speaker opts for an indirect mode of expression, such as responding to an invitation to a dinner party with “I don't have suitable clothes to wear”, comprehension is more demanding than simply saying “No, I won’t come”. This additional cost must be compensated for through richer contextual effects, which are typically interpreted as additional propositional inferences. Over the past two decades, experimental pragmatics has mainly focused on this information return, demonstrating that indirect refusals, ironic remarks, or metaphorical combinations do indeed generate supplementary implicit meanings [2][3][4]. However, natural conversations are filled with various indirect expressions, and their main benefits seem to lie in social, emotional, and even aesthetic aspects rather than merely information transmission. This kind of discourse behavior that requires additional reasoning steps can also demonstrate politeness, protect the dignity of the speaker, trigger a slight sense of humor, and showcase elegant literary grace. These non-verbal effects, sometimes mentioned in politeness studies [5] and in humor studies [6] , remain on the periphery of mainstream semantic theory frameworks. Therefore, we lack a systematic explanation of how different types of cognitive effects can jointly balance processing costs and jointly meet the principle of optimal relevance.

The present theoretical note aims to fill this gap and proposes a four-dimensional classification system for “additive effects”, which are the indirect manifestations that usually occur: interpersonal, humorous, aesthetic, and informational. Through publicly verified examples, it demonstrates that apparent irrelevance is often compensated for by multiple benefits that operate at different functional levels. The discussion emphasizes two testable hypotheses that can serve as the basis for future experimental research, and argues that if the concept of “contextual effect” of RT is to be able to explain why speakers frequently choose detours rather than straight paths, it must be decomposed.

2. Theoretical Gap

Despite decades of research on Relevance Theory (RT) and its contrast with default or neo-Gricean approaches, there is still relatively scattered empirical evidence regarding where the processing costs for indirect meaning arise. The experimental pragmatics research over the past decade has clarified

ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

some aspects of the relationship between effort and effect, which is central to RT, but has also revealed systematic blind spots that prompt us to re-examine this theory. A series of behavioral studies on the implicit meaning of quantifiers have shown that pragmatic richness does not occur automatically. Van Tiel, Pankratz and Sun demonstrated through a reaction time truth judgment task that readers experience measurable deceleration when deriving the inference "some → not all"[7]. The important point is that their handling of different scalar sets indicates that costs are context-dependent and not solely triggered by the uniformity of lexical forms. These findings partially support RT' prediction that reasoning effort will only be expended when the contextual benefits brought by reasoning are sufficient, but they do not clarify how this trade-off is calculated.

Electrophysiological evidence makes the situation even more complicated. Holtgraves and Kraus (2018) used event-related potentials to track the online comprehension of quantifier statements in conversations and found that for the interpretation of pragmatic meaning, there was an enhanced N400/P600 component compared to the interpretation of literal meaning [8]. Therefore, the temporal pattern of neural activity indicates that more integration efforts are required, but these ERP features cannot distinguish whether this cost is due to lexical retrieval, context update, or intention inference, and this distinction is of decisive significance in RT. Ronderos and Noveck recently conducted a study on process decomposition that directly addressed this issue [9]. They used pre-registered reaction time experiments and a large sample to distinguish between intention interpretation delay and mere decoding costs within the classical Bott & Noveck paradigm. Their research results confirm that the slowdown stems from the attribution of psychological states rather than sentence verification per se. This provides strong behavioral support for RT's claim that the calculation of relevance is a psychological interpretation calculation. However, even though this improved paradigm regards cognitive effort as a single indicator of delay, it has not quantified emotional or aesthetic effects.

Cross-linguistic and polarity-based evidence further enriches the content of the analysis. Van Tiel and Pankratz found that the polarity of adjectives affects the cost of implicit meaning[10]. For example, negative polarity adjectives (e.g., impolite → not very polite) e cause an increase in reading time, while positive polarity adjectives do not. The authors explain this asymmetry as a result of pragmatic expectations rather than a manifestation of lexical markers, which is consistent with the RT's effort–effect calculation model. However, their research and previous studies have not distinguished between the two different kinds of relevance that operate in actual discourse, namely the information motivation and the interpersonal motivation. Besides, the study of irony and developmental pragmatics has also revealed another gap. Olkoniemi et al. demonstrated through eye-tracking experiments that both children and adults would re-read sentences with ironic implications, indicating a delayed integration process and also reflecting emotional engagement [11]. Although these results confirm the predictions of RT, that inferring the speaker's attitude requires more reasoning steps, they also reveal the emotional returns resulting from this, namely humorous or interpersonal returns, which still exceed the scope of current quantitative models regarding relevance. After synthesizing these viewpoints, Khorsheed, Price and van Tiel reviewed the relevant literature and concluded that pragmatic costs arise from multiple interacting sources: semantic uncertainty, context update, and theory of mind calculation[12]. However, they also pointed out a persistent methodological limitation, that is, most experiments attribute heterogeneous cognitive rewards to a single reaction time or ERP amplitude, thereby burring RT's multidimensional concept of the "effect." Therefore, this field lacks a comprehensive framework that can map informative, social, and emotional returns to different processing features.

In sum, empirical evidence confirms that pragmatic richness requires measurable efforts, but the current paradigm defines the effort and effect too narrowly. RT predicts a broader mechanism that not only covers the accuracy of information but also includes interpersonal harmony and aesthetic appreciation. Therefore, the theoretical gap lies in linking these qualitatively different returns with quantitative indicators of cognitive processing. To fill this gap, experiments need to be designed to divide reaction times, neural activities, and emotional data into different motivational components, so that the balance between the effort and effect of reaction times can be tested as a truly multi-dimensional structure.

3. Research Implications

The empirical studies reviewed above all reached a core point: the effort of practicality is measurable, but its interpretation depends on the type of effect pursued. Yet from self-paced reading [7][10] to ERP[8] to eye-tracking[11], various paradigms compress various different motivations into a single delay or amplitude metric. This single curve covering all situations eliminates the multi-

ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

dimensional logic of RT. In Sperber and Wilson's formulation, processors would invest additional cognitive effort to achieve the expected effect of indirect discourse, whether the discourse is informative, interpersonal, or otherwise, and this additional cognitive effort is reasonable[1]. However, different effects will produce qualitatively different returns. Treating them as equivalent would distort the content that RT is intended to simulate.

Indirect expression can at least provide four distinguishable types of rewards. Firstly, informational effect. The hearer will obtain new semantic content that is consistent, e.g., the scalar inference[13]. RT predicts this as the typical case of cost–benefit calculation. Van Tiel et al. and van Tiel & Pankratz precisely simulated this information acquisition and found that context support regulates reaction time[7][10]. Secondly, interpersonal effect. The hearer infers the speaker's social stance or politeness level. None of the current research paradigm directly measure this reward, though Ronderos & Noveck's study indicates that the cost of interpreting intentions exceeds the magnitude delay, which is an implicit interpersonal factor[9]. Thirdly, humorous effect. When achieving entertainment effect through deliberate false means, the "cost" paid in this processing process may become beneficial. For satirical content, readers' re-reading time is longer, indicating this is a conscious re-processing process, and the effect is emotional satisfaction rather than information accuracy[11]. Fourthly, aesthetic effect. The hearer resonates with the style or metaphorical meaning of the sentence, such as in poetry or philosophical discourse. None of these six experiments quantified this reflective processing process, but RT explicitly licenses it as a poetic effect [1].

In various research paradigms, these dimensions are either not controlled or intentionally suppressed. Holtgraves & Kraus asked participants to judge truthfulness, excluding interpersonal interaction or humorous elements[8]. Van Tiel's tasks set an indication to accelerate verification, which ensures the clarity of reaction time data but removes irrelevant information associations. Even in in the improved design of Ronderos & Noveck's, participants treat utterances as propositions, not as jokes or social moves[9]. Consequently, the field has sufficient evidence for informational relevance, but there is almost no evidence on how other reward types regulate the effort threshold.

Khorsheed, Price & van Tiel (2022) pointed out this issue, stating that "cognitive cost" has been defined as a scalar variable, while "effect" still has multi-dimensionality and is partially emotional[12]. The response time theory predicts that the threshold for sufficient correlation should vary depending on the expected effect type. When the expected reward is social recognition or entertainment, hearers may tolerate or even seek more processing effort. However, when only factual accuracy is involved, the same delay may be considered overly lengthy. The averaging of these different motivations would produce a misleading monotonous "cost curve", masking the core economic principles of RT.

To make the theory match the methods, future experiments based on RT must shift from single-dimensional cost measurement to decomposition based on specific effects. Response time or ERP data should be accompanied by auxiliary indicators that can capture social, emotional, or aesthetic rewards, such as emotional self-evaluation, pupil dilation, or laughter response. Then, internal variable processing can be applied to the same stimulus, systematically changing the interpretive context (informational, interpersonal, humorous, or aesthetic) of the same stimulus to test whether the same sentence will produce different cost-reward characteristics.

It can be restated as an empirical research question: When the same language stimuli are placed in informative, interpersonal, humorous, or aesthetic contexts, will the processing cost pattern of indirect speech show systematic changes? And does RT not only predict the existence of these costs but also predict the gradient of their changes? To answer this question, it is necessary to decompose the overall delay into the driving factors of each component. This methodological shift will transform the reaction time test from a single-dimensional timing activity into a multi-dimensional correlation mapping. Such a framework will make the effort level in the experimental operation consistent with the initial definition of reaction time in terms of effect, thereby extending the theory from the accuracy of propositions to the full spectrum of human communicative reward.

4. Typology

Indirect speech is not a single pragmatic means; it encompasses multiple separable cognitive benefits. The phenomenon can be classified into four "effects": informativeness (new factual conditions), interpersonal (face preservation), humor (dissonance and release), and aesthetic (poetic taste). These effects, whether occurring alone or interacting, reveal their cognitive benefits. Each is defined, illustrated with a brief verified example (source attached), and followed by an outline of the

ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

theoretical gains. The final example how one utterance can activate several effects at once.

The first is informative effect: The declarative content that the hearer acquires is not presented through the language expression but based on the assumption that the speaker intends to provide as much information as possible according to the context requirements, which is a typical implicit meaning of magnitude. For example: "I ate some of the cookies." — a typical implicature example [13]. The surface meaning merely indicates that the speaker ate at least some cookies, but in the context involving the remaining quantity, the hearer usually infers that it is not all, thereby deriving the implicit meaning "some → not all." In relevance-theoretic terms, the hearer excludes unnecessary possible worlds ("the plate is empty"), thereby obtaining information accuracy. When the benefit of information acquisition exceeds the processing cost, this indirect inference is reasonable.

The second is interpersonal effect: The speaker uses an under-informative or attenuated literal form to mitigate face-threats, signal politeness or unity, or adopt strategic ambiguity about social rewards. For Example: "My, that's an interesting haircut." — film dialogue, The Three Stooges (2012) [14]. The literal meaning of the utterance is mildly positive, but in context (the addressee's unconventional appearance) it conveys polite criticism. From a RT viewpoint, the listener will tolerate higher uncertainty, or seek additional clues (intonation, facial expressions, etc.) to explain the expected attitude. Cognitive returns are social harmony, not factual information. Avoiding embarrassment and maintaining rapport justify the extra inferential effort.

The third is humorous effect: the speaker intentionally employs an under-informative or absurd utterance to create cognitive incongruity. It will be fun to solve the mismatch. For example: "I smell snow." — Lorelai Gilmore, Gilmore Girls (S1E08) [15]. The line is literally nonsensical, a person can't smell the snow, but in the cultural framework of the show, its function is Lorelai's iconic whimsy comment. The hearer must coordinate with the speaker's passion to produce humor and emotion. RT can satisfy this non-propositional reward so long as the "entertainment dividend" compensates for the additional processing cost.

The fourth is aesthetic effect: The speaker will obscure or blur the clear meaning to prompt the listener to think, appreciate aesthetically, or engage in poetic contemplation. This is precisely what RT refers to as the poetic effect. For Example: "Time is a flat circle." — Rust Cohle, True Detective (S1E05)[16]. This metaphor is literally false, but it can trigger thoughts about eternal repetition and fatalism. Its relevance does not lie in the narrow propositional content, but in the depth of reflection it evokes. When the expected aesthetic reward is high, the listener is willing to prolong the thinking process, repeatedly practice or re-read this line to obtain an emotionally rich experience beyond the propositional meaning.

A single turn can produce multiple effects. For example, when a character fails and says, "I killed it," the utterance may simultaneously convey informational, interpersonal, and humorous values: informational (indicating poor performance), interpersonal (maintaining self-image through understatement), and humorous (the incongruity between the literal meaning and the expected meaning). In reaction-time or eye-tracking data, these mixed actions are compressed into a single delay measurement value. Therefore, a multi-dimensional analysis is required, integrating behavioral, emotional, and social indicators to distinguish these different correlation calculations.

5. Discussion

RT stated that only when a series of sufficient contextual effects can prove that such efforts are reasonable will corresponding efforts be made[1]. The typology above indicates that the same phonological sequence can produce four different combinations of effects with distinct properties. Therefore, the processor is faced with a resource allocation problem: which effect type to expect, and how much cognitive effort to invest. This forces three specific revisions to the standard scalar-implicature of RT.

First, the traditional effort curve is no longer no longer a single indicator for measuring the difficulty; it is the visible tail of a distribution whose shape is determined by the weighted sum of expectations in the areas of information (I), interpersonal (P), humorous (H) and aesthetic (A) pay-offs. When only I is contextually plausible, the relevance metric collapses to the classic truth-conditional ratio: cognitive effect = eliminated worlds, effort = scalar-computation plus negation verification. The observed RT curve is therefore steep but short-lived, which is in line with the curve characteristics reported by Huang and Snedeker (2018), indicated that although numerical implicit meaning

ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

calculation is rapid, it is sensitive to predictability and intonation marking[17]. When P is added, the processor must simultaneously calculate the face-threat mitigation value. Due to the slower update speed of social-indexical variables (power, distance, affect) compared to the referential domain, the correlation threshold decreases, thereby allowing for deeper processing cycles and extending the right side of the reaction time curve, without necessarily increasing the mean; this explains why some polite sentences often produce a bimodal distribution, and once the social variables are fixed, this distribution disappears (cf. van Tiel & Pankratz 2022: Fig. 3)[10].

Second, the presence of H or A introduces a non-monetary cognitive currency: affective reward. Recent neuro-pragmatic evidence indicates that humorous solutions activate the reward system of the ventral striatum, resulting in dopamine responses related to pleasure and motivation[18]. In term of RT, this can be modeled as a negative-effort term: humorous effect = anticipated mirth reduces required cognitive effect. Thus, when Penny says "Yeah, I really killed. They just didn't know it yet."[19], the processor may voluntarily prolong ambiguity, manifested as sustained gaze time or delayed mouse initiation, laughter reward outweighs the additional milliseconds. This is contrary to the standard cost-sign: a longer reaction time is not an indication of difficulty, but rather a manifestation of strategic enjoyment, similar to the slow-motion replay in visual aesthetics. Combining such experiments with informative experiments will inevitably increase the differences, thereby creating an illusion of a unified processing cost.

Third, the multi-dimensional reward mechanism supports a parallel rather than serial derivation process. Unlike the classical two-step (literal → enriched), the processor can simultaneously activate multiple effect-based hypotheses and allocate activation levels based on the previous context (speaker identity, genre, medium). The observed RT then reflects the winner-takes-all collapse phenomenon rather than the sum of sequential steps. This provides a natural explanation for the "early-late effect" proposed by Olkoniemi et al.'s (2023)[11]: once the speaker's humorous style is mastered, the H hypothesis will win faster, although the same enrichment level is still calculated, the reaction time will decrease. The prior probability P (H| speaker) is strengthened, reducing the need for further evidence and shortening the decision-making time.

Collectively these revisions have transformed RT from a diagnostic of difficulty into a composite utility signal: RT = f (E [I, P, H, A] – C), where C is the classical Levinsonian computation cost and the expected term is influenced by the speaker, the style, and the trial history. The practical consequence is that any experimental comparison that does not orthogonalize the effect type is likely to confuse costs with voluntary input, thereby overestimating the "effortfulness" of pragmatic inference.

To evaluate the revised model, we proposed two pre-registered hypotheses that can be tested using existing button-press or eye-tracking set-ups. However, these hypotheses will explicitly change the type of effect while keeping the lexical form unchanged. Hypothesis 1 contrasts informational vs interpersonal effects within-subjects. Using the Ripley cookie line[13], one block instructs participants to "decide how many items are left" (Information), the other presents the same sentence spoken by a crew-member rebuked for eating rations and asks participants to "rate how polite the speaker is" (Face). Visual arrays are identical (e.g. 4 cookies on screen). Multidimensional interpretation predicts: (a) scalar-implicature rates will be equivalent (both require 'not all'); (b) RT will be longer in the Face block only for participants high in interpersonal sensitivity (measured by Davis Empathy Scale); (c) diffusion-model drift rates will show delayed collapse of evidence in the Face block, indicating a deeper rather than more difficult processing process. The current dataset cannot meet the orthogonal operation conditions; therefore, this prediction is indeed forward-looking. Hypothesis 2 traces a humor-decay curve. Penny's audience line[19], is embedded in a rapid-serial humour paradigm: 20% of the trials were jokes, and 80% were declarative sentences. Eye-tracking measurement (the time spent gazing at the speaker's mouth) can be used as an indicator of autonomous maintenance behavior. This model predicts that at the N positions from the beginning to the end of the joke, the gaze time will monotonically decrease, while the accuracy of implicit meaning remains stable. This indicates that the processor has learned to reduce the allocation of enjoyment time while still obtaining rich experiences. Crucially, if the classical cost perspective is correct, then accuracy should increase with the increase in speed; while from the multi-dimensional perspective, accuracy remains unchanged because the richness of experience itself does not require effort, but only the optional enjoyment time is shortened.

These assumptions can be simply applied to any laboratory that has already run visual world or autonomous reading experimental models; no customized hardware equipment is required. However, this paper does not provide new empirical evidence. All the arguments are re-analyses of the published overall reaction time data, and the original design of these data led to confusion in the type of effect. Before the orthogonal operation mentioned above, the multi-dimensional model is still based on the

ISSN 2616-5783 Vol.8, Issue 12: 23-28, DOI: 10.25236/AJHSS.2025.081204

inference of the existing variance model rather than an improved version derived from the data.

6. Conclusion

This note decomposes the simple "scalar implicit cost" into four explanatory effects: informational, interpersonal, humorous, and aesthetic, and demonstrates that each effect re-adjusts the "effort-reward" balance of RT, thereby predicting that a longer reaction time implies deeper calculation, reduced social risk, voluntary taste experience, or acquired expectations. By mapping the existing delay margin onto this multi-dimensional grid, this theory resolves the previous contradictions while not abandoning the core cognitive insight of reaction time. The typology provides a principled explanatory approach for second language teachers and conversation analysis experts to explain why the same discourse feels easy to understand in a joke but is extremely difficult in a laboratory declarative task: the classroom must first indicate the expected effect type before requiring learners to invest cognitive resources. Therefore, material designers can shorten the teaching time by highlighting interpersonal or humorous elements rather than purely declarative values. The proposed orthogonal operation method prompts experimental pragmatics to shift from monotonous repetitive experimental competitions to the precise modeling of why speakers from different language backgrounds and cultural backgrounds are willing to pay the "privilege cost" of not directly expressing their true thoughts.

References

[1] Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition. Oxford: Blackwell.
[2] Gibbs, R. W. (1999). Intentions in the experience of meaning. Cambridge University Press.
[3] Noveck, I. A. (2018). Experimental pragmatics: The making of a cognitive science. Cambridge University Press.
[4] Cummins, C., & Katsos, N. (Eds.). (2019). The Oxford handbook of experimental semantics and pragmatics. Oxford University Press.
[5] Brown, P., & Levinson, S. C. (1987). Politeness: Some universals in language usage. Cambridge University Press.
[6] Attardo, S. (2020). The linguistics of humor: An introduction. Oxford University Press.
[7] Van Tiel, B., Pankratz, E., & Sun, C. (2019). Scales and scalarity: Processing scalar inferences. Journal of Memory and Language, 105, 93–107.
[8] Holtgraves, T., & Kraus, B. (2018). Processing scalar implicatures in conversational contexts: An ERP study. Journal of Neurolinguistics, 46, 93–108.
[9] Ronderos, C. R., & Noveck, I. A. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. Cognition, 238, 105480.
[10] Van Tiel, B., & Pankratz, E. (2022). Adjectival polarity and the processing of scalar inferences. Glossa: A Journal of General Linguistics, 6(1), 32.
[11] Olkoniemi, H., Halonen, S., Pexman, P. M., & Häikiö, T. (2023). Children's processing of written irony: An eye-tracking study. Cognition, 238, 105508.
[12] Khorsheed, A., Price, J., & van Tiel, B. (2022). Sources of cognitive cost in scalar implicature processing: A review. Frontiers in Communication, 7, 990044.
[13] Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. MIT Press.
[14] Farrelly, P., & Farrelly, B. (Directors). (2012). The Three Stooges [Film]. Twentieth Century Fox. Dialogue retrieved from Clip.cafe (verified film transcript).
[15] Sherman-Palladino, A. (Creator). (2000). Gilmore Girls [TV series]. Episode "Love and War and Snow," Season 1, Episode 8. Warner Bros. Television. Transcript available at Subslikescript.com.
[16] Pizzolatto, N. (Writer), & Fukunaga, C. J. (Director). (2014). True Detective [TV series]. Episode "The Secret Fate of All Life," Season 1, Episode 5. HBO. Transcript available at Subslikescript.com)
[17] Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. Cognitive Psychology, 102, 105–126.
[18] Prenger, M., Gilchrist, M., Hedger, K., Seergobin, K., Owen, A., MacDonald, P. (2023). Establishing the roles of the dorsal and ventral striatum in humor comprehension and appreciation with fMRI. The Journal of Neuroscience, 43(49), 8536–8549.
[19] Lorre, C. (Creator), & Prady, B. (Creator). (2008). The Big Bang Theory [TV series]. Episode "The Barbarian Sublimation," Season 2, Episode 3. Warner Bros. Television / CBS. Transcript available at BigBangTrans or Subslikescript.com.