Sociology, Philosophy and Psychology; Vol.3, No.2; 2026
E-ISSN: 3006-3620 P-ISSN: 3006-3612
DOI: https://doi.org/10.70267/spp.v3n2.5669
Published by: Zeus Press

The Cognitive Origins of Individual Differences in Pragmatic Competence: A Theoretical Review and Outlook from the Perspective of Predictive Coding

Luna Yang*

School of Education, Shanghai Normal University TIANHUA College, Shanghai 201815, China

**Corresponding author: Luna Yang.*

Abstract

Pragmatic reasoning ability is essential to interpersonal communication, and individuals differ markedly in it, yet its cognitive origins remain unclear. In a large-scale study of individual differences, Floyd and colleagues showed that pragmatic language use comprises three relatively independent components: social conventions, intonation, and world knowledge–based causal reasoning. Building on this work, the present paper adopts a predictive coding perspective and proposes a theoretical framework. Its central claim is that individual differences in three stages of predictive coding, namely prediction generation, prediction error encoding together with precision weighting, and adaptive updating of predictive models, may be a common cognitive source of the individual differences observed across the three components of pragmatic reasoning. Specifically, the social conventions component may correspond primarily to the generation and precision regulation of high-level social priors; the intonation component to the precision weighting of low-level sensory input; and the causal reasoning component to the structure and content of world knowledge priors. The framework also offers a tentative explanation of why the three components are both correlated and separable: covariation may stem from a shared predictive coding infrastructure, whereas separation may stem from domain-specific prior content and stage-specific processing bottlenecks. The paper closes with several directions for future research. As a theoretical review, its core hypothesis remains to be tested by future empirical work.

Keywords

predictive coding, pragmatic reasoning ability, individual differences, three components of pragmatic ability, theoretical review

1. Introduction

Intentionality is a defining feature of linguistic communication. In everyday interaction, speaking is not merely the production of sound. We exchange information, signal our intentions, thoughts and stances, and sometimes convey meaning that is never stated outright. Understanding what is left unsaid is a core social skill. It shapes how smoothly workplace communication runs and how well relationships are maintained. Grasping the meaning behind a speaker's irony, indirect requests, or polite refusals all depends on this ability. This capacity to combine the immediate situation with shared experience in order to work out what a speaker really means is known as pragmatic reasoning, and it varies substantially across individuals [1]. Consider a manager who says, "This report is fine as it is." Some listeners pick up the unstated message: basically acceptable, but

© 2026 The Author(s). Authors retain copyright in their articles.

lacking in depth and polish, and in need of further work. Others miss it entirely, and may even lose a promotion as a result. This raises a key question: what are the cognitive origins of these individual differences?

Previous work has advanced along two lines, one concerned with structure and the other with mechanism. On the structural side, Floyd and colleagues [2] combined a large-scale individual differences study (n = 776) with exploratory factor analysis. Their results showed that pragmatic language use is not a single ability. It consists of three relatively independent components: understanding social conventions, interpreting intonation patterns, and drawing causal inferences from world knowledge. Because every task in that study tapped comprehension rather than production, the present paper refers to the three components together as the "three components of pragmatic reasoning ability", in order to mark their status as the internal structure of pragmatic reasoning. This term is used consistently throughout. Although the study provides empirical grounding for the structure of pragmatic reasoning, it does not explain why the three components show stable individual differences. On the mechanistic side, predictive coding theory [3] holds that the brain is not a passive receiver of information. It continuously generates predictions, compares them against actual input, and revises them when a prediction error occurs. Language comprehension depends on this ongoing prediction, and the mechanism has been demonstrated at the lexical, syntactic and prosodic levels [4]. More importantly, predictive coding ability itself varies between individuals. Reaction times on Cloze tasks correlate with N400 amplitude, and the strength of that correlation also varies across individuals, which indicates that predictive coding differs from person to person [5].

Two bodies of research have thus been described: the structural study of pragmatic reasoning, and the predictive coding account of language comprehension. Each has advanced on its own, but the two have not been connected, and this leaves two gaps that need to be filled. First, although Floyd and colleagues [2] revealed the components of pragmatic language use, they did not explain where the individual differences in those components come from. A structural account can tell us what pragmatic reasoning is made of, but not why some people are better at it than others. Second, although predictive coding research has established that the brain is a continuous prediction machine, and that predictive coding ability varies across individuals, there is no direct evidence that this mechanistic variation is the cognitive source of individual differences in pragmatic reasoning. In other words, predictive coding offers an explanatory tool, but few studies have used it to explain individual differences in pragmatics.

In response to these gaps, earlier work has approached the question from social cognition, for example theory of mind, and from language experience, for example vocabulary knowledge and working memory [1]. These perspectives, however, tend to explain one component or one group at a time. Theory of mind is linked mainly to the social conventions component [2], and language experience mainly to general pragmatic performance [1]. Neither can serve as a unifying underlying logic. Neither explains why individual differences in the three components are systematic and stable, and neither accounts for the pattern in which the three components are both correlated and separable [2]. The present paper therefore advances a core hypothesis: individual differences in predictive coding are a common cognitive source of the individual differences found in the three components of pragmatic reasoning.

Specifically, the three components of pragmatic reasoning can be mapped onto three parametric stages of the predictive coding framework [3]. The social conventions component corresponds mainly to the generation and precision regulation of high-level social priors; the intonation component mainly to the precision weighting of low-level sensory input; and the causal reasoning component mainly to the structure and content of world knowledge priors. This mapping is proposed as a theoretical hypothesis. Its aim is to build a testable bridge between the structure of pragmatic reasoning and the mechanisms of language comprehension.

One caveat is in order: overall pragmatic performance is systematically related to an individual's neurodevelopmental status [6]. To delimit the scope of this review, neurotypical individuals are defined here as healthy individuals with no clinical history of autism spectrum disorder, attention deficit hyperactivity disorder, or other neurodevelopmental disorders, and no marked abnormalities in social cognition reported in previous research. On this definition, the pragmatic performance of neurotypical individuals serves as the reference point for what is called the "baseline prior". The pragmatic profile of non-typical groups, such as individuals with ASD, is introduced only as supporting evidence when the boundary conditions of the predictive coding model are discussed [7]. It should also be noted that this is a review, and no single social cognition scale is adopted as the criterion for defining the groups. How the "normal range" is operationalized varies from study to study.

© 2026 The Author(s). Authors retain copyright in their articles.

The rest of the paper is organised as follows. Chapter 2 defines the core concepts and sets out the theoretical basis of predictive coding. Chapter 3 examines the possible cognitive origins of individual differences in pragmatic reasoning, focusing on three stages: prediction generation, error encoding and precision weighting, and model updating. Chapter 4 proposes an integrative theoretical framework that offers a tentative account of why the three components identified by Floyd and colleagues [2] are both correlated and separable. Chapter 5 sets out testable hypotheses and directions for future research, and Chapter 6 concludes.

2. Core Concepts and Theoretical Basis

Before turning to the specific definitions, it is worth clarifying which version of predictive coding theory this paper adopts. Predictive coding is not a single theory but a family of theories. The branches share a core mechanism of prediction, error and updating, but they differ in emphasis and in the range of phenomena they aim to explain. This paper adopts Clark's [3] Predictive Processing framework, because its level of analysis is closest to the cognitive processes of pragmatic reasoning that concern us here. That formulation also specifies how prediction generation, precision weighting and model updating interact dynamically, which gives us a direct theoretical tool for mapping the three components of pragmatic reasoning onto different stages of predictive coding in Chapter 3. Clark's framework is also influential across cognitive science. It has been applied to language processing, social cognition and clinical psychology, and so can support the cross-disciplinary integration attempted in this paper.

Where other versions of predictive coding are cited, such as the HIPPEA hypothesis of Van de Cruys and colleagues [6], they are treated as extensions and refinements of Clark's [3] framework, so that the framework retains its central position.

2.1 Concept Definition and Composition of Pragmatic Inference Ability

This paper uses the term pragmatic reasoning ability to refer to the capacity to infer a speaker's implicit communicative intent in real-time interaction, on the basis of both the literal meaning of an utterance and information from the context. Pragmatic competence usually covers both production and comprehension. Pragmatic reasoning, by contrast, focuses on the dynamic processes involved when that competence is deployed in real time. Pragmatic reasoning ability is therefore the central object of study here, and the three components of pragmatic reasoning ability are treated as its internal constituents within the framework proposed by Floyd and colleagues [2]. The theoretical roots of this ability lie in Grice's [8] cooperative principle and theory of conversational implicature. Grice [8] argued that successful communication does not rely on the literal encoding of linguistic symbols alone. Both parties must draw on a set of implicit cooperative conventions, such as the maxims of quantity, quality, relation and manner, in order to work out what the speaker means. The idea was subsequently developed as Levinson and colleagues [9] systematised the inferential mechanisms involved.

Is pragmatic reasoning a single, indivisible ability, or is it made up of several separable components? For a long time, this question lacked a large-scale empirical test. Floyd and colleagues [2] provided the first empirical account of the internal structure of pragmatic language use, in a large-scale individual differences study of 776 participants. Participants completed 21 tasks covering 11 pragmatic phenomena, and their performance was subjected to exploratory factor analysis. Once general cognitive ability was controlled for, the analysis consistently revealed three relatively independent components, shown in Table 1.

Table 1: The three components of pragmatic language use identified by Floyd and colleagues [2]

Component	Phenomena covered	Core cognitive operations
Social conventions	Indirect requests, conversational implicatures, irony	Understanding and following the face-saving conventions of social interaction, so that criticism, requests or evaluations can be conveyed through indirect, softened, or even literally opposite expressions
Intonation	Contrastive prosody, emotional prosody	Extracting information structure or emotional state from variation in the acoustic pattern (pitch, duration, stress)
Causal reasoning	World knowledge-based causal inference	Drawing on knowledge of causal relations between events in order to infer the logical connections between utterances

© 2026 The Author(s). Authors retain copyright in their articles.

The "core cognitive operations" column of Table 1 summarises the main cognitive processes on which each component depends. Chapter 3 links each of these processes to a different stage of the predictive coding framework. The theoretical resources that the three components draw on, such as politeness theory and semantic memory networks, are discussed in the relevant sections.

2.2 Behavioral Manifestations of Individual Differences in Pragmatic Inference

If pragmatic reasoning ability is a static store of capacity, pragmatic reasoning is the dynamic process through which that capacity operates in real communication. Individual differences in pragmatic reasoning can be quantified in behavioural experiments using measures such as reaction time (RT), accuracy, and eye movement patterns.

In the processing of indirect requests, Marocchini and colleagues [10] found that individuals with stronger theory of mind (ToM) ability read the critical region "Can you" in conventionalised indirect requests such as "Can you...?" significantly faster than individuals with weaker ToM ability, who needed more time. This indicates that the speed with which people recognise the intent behind an indirect request varies in measurable ways, and that this variation is closely tied to social cognitive ability.

In intonation interpretation, Esteve-Gibert and colleagues [11] found that listeners with higher empathy use prosodic cues to form predictions before the lexical information has been fully presented. Listeners with lower empathy process prosodic meaning less effectively, and have to wait for subsequent lexical information before reaching a judgement. This shows that differences in how efficiently people extract pragmatic information from acoustic signals can be quantified using eye movement measures.

Taken together, these two studies show that individual differences in pragmatic reasoning are measurable at the behavioural level, and that they are associated with stable social cognitive traits such as theory of mind and empathy. Why some people can rapidly draw on social cognitive resources to process an indirect request, while others need more time, is the cognitive question behind these differences, and it is the question this paper tries to answer from a predictive coding perspective.

2.3 Core Mechanism of Predictive Coding Theory

Predictive coding is one of the dominant frameworks in contemporary cognitive neuroscience for explaining how the brain works. In his classic formulation, Clark [3] argued that the brain is not an input-output device that passively receives sensory information. It is a prediction machine: it continually generates predictions, matches incoming sensory signals against top-down expectations, and updates its internal models in response to prediction error. The core claims of the theory can be broken down into the following key concepts.

First, the hierarchical generative model. The brain contains a multi-level generative model. Higher levels represent abstract information, such as context and intention, while lower levels represent concrete information, such as acoustic features. Higher levels send prediction signals downward, and lower levels send error signals upward, reporting the gap between what was expected and what actually arrived.

Second, prediction error. A prediction error arises when a high-level prediction conflicts with the actual input at a lower level. This error is the core signal that drives learning, and the brain minimises it continuously by adjusting its predictions or its actions.

Third, precision weighting. Not all prediction errors carry the same weight. The brain estimates the reliability of each error signal. Errors with high precision are given greater weight and are more likely to trigger a model update, whereas errors with low precision are relatively ignored.

Fourth, adaptive updating of the predictive model. When a weighted error is judged to be reliable and meaningful, the brain passes it up the hierarchy, prompting higher-level systems to adjust the parameters of the generative model. This update allows the brain to learn from prediction failure and to refine its internal models over time.

In language processing, predictive coding has been shown to operate at multiple levels. Listeners predict upcoming words at the lexical level and upcoming structures at the syntactic level [4]. Most of this work, however, focuses on the general predictive mechanisms of language processing. A few studies have begun to

© 2026 The Author(s). Authors retain copyright in their articles.

examine individual differences in predictive coding, for example Jano [5], but no systematic theoretical link has been drawn between these differences and individual differences in pragmatic reasoning.

2.4 Preliminary Hypotheses of Theoretical Correlation

Sections 2.1 to 2.3 lead naturally to a question: do the three components of pragmatic language use identified by Floyd and colleagues [2] correspond to individual differences in different stages, or different parameters, of the predictive coding framework?

A preliminary mapping can be stated as follows.

The core cognitive operation underlying the social conventions component is inferring the speaker's implicit intent and following the social conventions of "face". At heart, this is a high-level prediction about other people's mental states and about social convention. Individual differences in this component may therefore arise mainly from differences in the richness and precision of social priors, corresponding to the prediction generation stage, and in the precision weighting given to high-level priors, corresponding to the precision weighting stage.

The core cognitive operation underlying the intonation component is extracting pragmatic information from acoustic patterns. At heart, this is a low-level prediction about sensory input, together with an assessment of its precision. Individual differences in this component may therefore arise mainly from differences in the precision weighting given to sensory input, again at the precision weighting stage.

The core cognitive operation underlying the causal reasoning component is drawing on knowledge of causal relations between events. At heart, this is a call on the content of the generative model of how the world works. Individual differences in this component may therefore arise mainly from differences in the richness and structure of the causal knowledge network within the generative model, corresponding to the prediction generation stage.

These preliminary hypotheses are developed systematically in Chapter 3, which examines how individual differences are expressed in each of the three stages of predictive coding, and argues that together these differences shape the overall pattern of individual differences across the three components of pragmatic reasoning.

3. Mechanism of Individual Differences in Pragmatic Competence from the Perspective of Predictive Coding

3.1 Individual Differences in Predictive Generation: How Prior Beliefs Shape the Starting Point of Pragmatic Reasoning

3.1.1 Differences in the Richness and Precision of Social Priors

According to predictive coding, the brain continually generates top-down expectations about upcoming sensory input, and the content of those expectations depends to a large extent on the prior knowledge the individual already has [3]. In pragmatic processing, this expectation takes the form of a listener's anticipation, based on the current context and past experience, of what the speaker is about to mean. Individuals differ at this anticipatory stage, and one possible reason is that they differ in the richness and precision of the social priors they hold.

The social conventions component identified by Floyd and colleagues [2] covers prototypical pragmatic phenomena such as indirect requests, conversational implicatures and irony. Processing these phenomena depends heavily on how well the listener has mastered the conventions of social interaction and the strategies available for saving face. For example, on hearing "This report is fine as it is", efficient reasoners quickly draw on social schemas about workplace hierarchy and about indirect criticism, and form the high-level expectation that their supervisor may not in fact be satisfied. When the literal meaning of the utterance conflicts with that expectation, the ironic intent can be recognised. Less efficient reasoners, by comparison, may lack well-developed links in their social schemas between refusing indirectly and leaving room for improvement. They therefore find it harder to generate the high-level expectation, and are more likely to stay at the literal level.

© 2026 The Author(s). Authors retain copyright in their articles.

From a predictive coding perspective, one possible explanation of this difference lies in variation in the precision of social priors. Under precision weighting, when an individual assigns high precision to their social priors, those priors exert a stronger influence on how subsequent linguistic input is interpreted. Conversely, if the precision of social priors is low, the individual will find it harder to move beyond literal meaning when the pragmatic input is ambiguous, and pragmatic reasoning suffers. The stable individual differences seen in the social conventions component may therefore be explained in terms of individual specificity in the precision of high-level priors. This hypothesis, of course, awaits direct empirical test.

3.1.2 Structural Differences in World Knowledge Priors

Beyond social priors, pragmatic reasoning also draws on another important prior: knowledge of causal relations in the objective world, or the world knowledge prior. The third component identified by Floyd and colleagues [2], world knowledge–based causal reasoning, is a measure of exactly this kind of prior. Its tasks ask participants to judge whether two events are causally related, for example whether "the cat ran around the room" is related to "the vase was broken" [2].

From a predictive coding perspective, how efficiently a person performs causal reasoning depends on how rich and finely detailed their representation of event causal structure is within the generative model. Because of differences in upbringing, education, professional training and everyday experience, the causal knowledge networks of different individuals differ systematically in both breadth and strength of connection. Someone with a medical background, for instance, can quickly generate the causal prediction "this may be a drug allergy" when told that a patient developed a rash after taking a new medication. Someone without that background struggles to establish the causal link, and pragmatic reasoning becomes slower or goes astray.

Notably, the separation between world knowledge priors and social priors [2] maps neatly onto two kinds of prior knowledge that sit at different levels and in different domains within the predictive coding framework. Social priors concern high-level inference about intentions, beliefs and social conventions, whereas world knowledge priors rest more on long-term experience of physical causation and event logic. Structural differences in world knowledge priors may therefore be an important source, at the prediction generation stage, of the individual differences seen in the causal reasoning component.

3.1.3 Differences in "Pragmatic Sensitivity" of Predictive Generation

Sections 3.1.1 and 3.1.2 dealt with differences in the content of priors: what knowledge an individual has, and how precise that knowledge is. Section 3.1.3 asks a different question. Suppose two people hold equally rich and equally precise priors. Will their pragmatic reasoning be identical? What is at issue here is how willing a person is to actively deploy those priors when generating pragmatic predictions.

This dimension is called pragmatic sensitivity. It refers to how aware an individual is of pragmatic cues in the communicative situation, such as the speaker's identity, the tone of the context and the degree of cooperation in the exchange, and how inclined they are to use those cues to generate predictions about the speaker's intent [1].

Even when two people have similar stores of prior knowledge, their pragmatic performance in real-time interaction can still differ markedly, and the reason lies in individual differences in two sub-processes.

The first is differences in the threshold for noticing pragmatic cues. In conversation, speakers rarely state their intent outright. They convey it through subtle linguistic markers, such as particles and sentence-type choices, or through contextual cues, such as where an utterance falls in the sequence and when the topic shifts. Individuals differ in how sensitive they are to these "pragmatic triggers". Some pick up immediately on a manager's sudden use of an informal form of address in a formal setting, and quickly predict that he may be trying to build rapport. Others miss the signal altogether and carry on with their original pragmatic expectations [2].

The second is differences in the motivation to deploy priors. Noticing a pragmatic cue is only the first step. The individual then makes an implicit decision about whether to invest cognitive resources in retrieving the relevant priors and generating a prediction about intent. This deployment is not automatic. It is modulated by communicative motivation and by engagement tendency. When motivation is high, as in a conversation with a superior or in a situation where cooperation matters, people are more willing to invest effort and actively

© 2026 The Author(s). Authors retain copyright in their articles.

deploy their priors. When motivation is low, as in casual chat, they fall back on the least effortful route, a literal reading, even when the relevant priors are available to them.

From a predictive coding perspective, differences in pragmatic sensitivity reflect the fact that prediction generation is conditional. Prior knowledge is not a static tool that processes information automatically. Its deployment is a dynamic process shaped jointly by the situation, the individual's state and their motivation. This may partly explain why the same person performs at very different levels in different contexts: the threshold for deploying priors, and the motivation to do so, are activated to different degrees in different situations [1].

Pragmatic sensitivity also interacts, in principle, with the precision of social priors discussed in Section 3.1.1. An individual with low pragmatic sensitivity may hold highly precise social priors and still fail to deploy them, simply because the relevant cues were never noticed. The result is someone who has the ability but does not perform well. This distinction matters for the intervention and training discussion in Chapter 5. If a person's pragmatic difficulty stems from a failure to deploy priors rather than from a lack of priors, then training that only enriches the priors will have limited effect. It has to be combined with ways of raising sensitivity to pragmatic cues and the motivation to act on them.

3.2 Individual Differences in Prediction Error Coding and Precision Weighting

3.2.1 Differences in Neural Efficiency of Error Coding

Once a prediction has been generated, the next critical stage is the encoding and processing of prediction error [3]. In pragmatic reasoning, this stage determines whether the listener can detect the conflict signal that arises when literal meaning and context do not match, and so whether corrective processing is triggered.

Prediction error in language processing has a clear neural signature: the N400 component. When a person encounters a linguistic stimulus that does not fit their expectations, N400 amplitude increases significantly [12]. Within a predictive coding framework, N400 amplitude can be read as the strength of the prediction error signal: the larger the error and the more unexpected the input, the larger the amplitude [13]. Crucially, both the amplitude and the latency of the N400 vary systematically across individuals [4]. Some people show faster and larger responses to pragmatic prediction errors, which indicates greater sensitivity to error detection. Others respond more slowly or more weakly, indicating lower efficiency in error encoding.

Extending this logic to the pragmatic level yields the following hypothesis. In phenomena such as irony and indirect requests, where interpretation has to go beyond the literal, the listener must first detect the prediction error that arises when the literal reading conflicts with the context. If neural encoding is efficient, this conflict signal is picked up quickly and corrective processing is triggered. If encoding is inefficient, the conflict signal is not fully represented, the listener remains at the literal level, and the outcome is pragmatic difficulty [7]. Individual differences in the efficiency of prediction error encoding, as indexed by the N400, may therefore be a second key route to explaining individual differences in pragmatic reasoning.

3.2.2 Individual Differences in Precision Weighting and "Literalism"

Once a prediction error has been detected, the brain faces a critical computational decision: how trustworthy is this error signal, and how much weight should it carry in updating the internal model? In the predictive coding framework, this decision is made by the precision mechanism. Precision is essentially an estimate of the signal-to-noise ratio. High-precision signals are given greater behavioural weight, while low-precision signals are relatively discounted [3].

The importance of precision weighting in pragmatic reasoning comes into focus in a central dilemma. When a listener hears an utterance whose literal meaning clearly clashes with the context, as in "Lovely weather today" said during a downpour, the brain receives two conflicting signals at once: a high-level social prior, namely that nobody says something obviously false for no reason, and a low-level literal semantic input, namely that he said the weather is good. The listener has to weigh the estimated precision of the two signals and decide: trust the literal input and discount the social prior, which favours a literal reading, or trust the social prior and treat the literal input as an unimportant error signal, which favours an ironic reading. The outcome of this weighting determines whether pragmatic reasoning succeeds.

© 2026 The Author(s). Authors retain copyright in their articles.

Individual differences at this stage can be understood at two levels: in neurotypical populations and in atypical populations. Among neurotypical individuals, estimates of the precision of contextual cues and social priors vary systematically. When the pragmatic input is ambiguous, some people tend to assign high precision weight to social priors. They quickly discount the literal semantic input and arrive at a non-literal interpretation. Others assign excessive precision to the literal semantic input. The social prior is then relatively down-weighted, and reasoning becomes hesitant or fails outright. These individual differences in precision weighting may be related to general cognitive resources such as working memory capacity and cognitive flexibility [1].

Among atypical populations, the hypothesis of abnormal precision has been developed more systematically. Van de Cruys and colleagues [6] proposed the hypothesis of High, Inflexible Precision of Prediction Errors in Autism, or HIPPEA. On this view, individuals with ASD assign uniform and inflexible precision weights to prediction errors at the neural processing level, rather than adjusting them flexibly according to the situation. The direct consequence is that, when facing pragmatic phenomena such as irony and indirect requests, where the literal meaning must be partly set aside in favour of a deeper reading of intent, literal semantic input is treated as a sensory prediction error and given excessive precision. It is not easily suppressed by social priors, and the result is a tendency towards literalism. It should be noted that recent work suggests the "high precision" component of HIPPEA has not received consistent empirical support, whereas the "inflexibility" component has been supported more widely.

Vicente and colleagues [7] extended this reasoning to language processing and offered a more precise mechanistic account. In autistic language comprehension, the balance between prediction and error signals is abnormal, which leaves individuals with little confidence in the predictions they generate. Faced with non-literal language, this uncertainty leads them to fall back on the interpretation they trust most: the literal meaning, that is, the sensory prediction. To sum up, individual differences in precision weighting, whether the flexible-adjustment differences seen in neurotypical populations or the high and inflexible profile found in ASD, constitute yet another key route to explaining individual differences in pragmatic reasoning. This route complements the neural efficiency of error encoding discussed in Section 3.2.1. The latter concerns whether an error can be detected at all; the former concerns how much weight it is given once detected.

3.3 Individual Differences in Adaptive Updating of Predictive Models: Cognitive Origins of Pragmatic Flexibility

The first two stages of the predictive coding cycle, prediction generation (3.1) and prediction error encoding (3.2), together answer the question of how accurate a person's pragmatic predictions are in a given situation. But pragmatic reasoning is also fundamentally dynamic. Speakers move between contexts and between interlocutors, and their pragmatic habits change over time. An individual who copes well with this change has an advantage that lies not only in the accuracy of any single prediction, but in how quickly and how effectively they learn from prediction errors and update their internal models. This capacity is adaptive updating of the predictive model, and it is the function of the third core stage of the predictive coding framework.

The basic logic of model updating is as follows. When a prediction error signal has been precision weighted and judged to be reliable and meaningful, the brain propagates it up the hierarchy, so that higher-level systems can adjust the parameter structure of the generative model [3]. At the pragmatic level, this means that when a listener's literal prediction does not match the actual input, as when "lovely weather" is said during a downpour, the error signal must travel up to the level of intention inference and trigger a reassessment of what the speaker might mean. Individual differences in the efficiency of this propagation and updating process are the core cognitive source of pragmatic flexibility.

Differences in the speed of model updating may arise at several levels. The first is the threshold for updating. Some individuals need only a small or weak prediction error to trigger an update. These are the sensitive pragmatists: they notice the slightest mismatch between what was said and the context, and adjust their predictions for what follows. Others require an accumulation of large or strong error signals before updating. These are the insensitive pragmatists: even when the other person has signalled something quite clearly, they hold on to their initial prediction. The second is the depth of updating. Some individuals complete a cross-level adjustment quickly, moving from low-level literal meaning to high-level intention. Others keep their updating at the lower level, concluding perhaps that they misheard a word, without rising to the pragmatic level and asking whether the speaker meant something else entirely. This difference is likely to be closely tied

© 2026 The Author(s). Authors retain copyright in their articles.

to executive functions such as cognitive control and working memory capacity [1]. Individuals with ample cognitive control resources are better able to redirect low-level error signals to the higher-level pragmatic reasoning system, and so to achieve cross-level model adjustment.

It is worth emphasising that model updating is conceptually distinct from the precision weighting described in Section 3.2.2. Precision weighting determines whether an error signal is worth using for an update. Model updating determines how quickly and how deeply that update is carried out once the decision has been made. Together, the two determine how adaptive and flexible pragmatic reasoning is in dynamic communicative situations.

To sum up, individual differences in adaptive model updating, whether in the speed of updating during real-time conversation or in the structural shaping that comes with long experience, may constitute a third core route to explaining individual differences in pragmatic reasoning. Together with the first two stages, prediction generation and error encoding, it points to a hypothesis that remains to be tested: individual differences across the three stages of predictive coding jointly constitute the complete cognitive mechanism underlying individual differences in pragmatic reasoning.

4. Theoretical Integration: How the Three Links of Predictive Coding Jointly Shape Individual Differences in Pragmatic Inference Ability

The preceding chapters examined the possible cognitive origins of individual differences in pragmatic reasoning one stage at a time: prediction generation (3.1), prediction error encoding and precision weighting (3.2), and adaptive updating of predictive models (3.3). In real pragmatic processing, however, these three stages are not modules that operate in isolation and in sequence. They form a continuous, bidirectionally interacting predictive loop. High-level priors drive prediction generation from above, low-level errors drive model updating from below, and precision weighting dynamically balances the top-down and bottom-up signals at every level and at every moment [3]. Discussing the three stages separately brings analytical clarity, but it risks obscuring their interactions and their dynamic character. This chapter integrates the framework. It argues that the three stages of predictive coding jointly provide a systematic explanation of individual differences in pragmatic reasoning, and it returns to the question raised at the outset: why do the three components identified by Floyd and colleagues [2] show a pattern that is both correlated and separable?

4.1 Dynamic Interaction of the Three Links: A Self-Reinforcing Predictive Cycle

As Chapter 3 described, pragmatic reasoning begins when the listener draws on prior knowledge, both social and world knowledge, to generate a prediction about the speaker's intent (3.1). When the actual linguistic input does not match the prediction, the brain generates a prediction error signal, whose neural signature can be seen in amplitude changes in ERP components such as the N400 (3.2.1). The error signal is then precision weighted, so that the brain can judge whether it is worth using to update the internal system (3.2.2). Once it is judged valid, it is propagated up the hierarchy and triggers parameter adjustments in the generative model at higher levels (3.3).

These three stages, however, do not simply operate in isolation. They form a cyclic system with three kinds of interaction.

Interaction 1: prior precision and prediction generation shape each other. The precision of an individual's priors in a given domain, the core parameter of precision weighting (3.2.2), directly determines how strong and how confident their predictions in that domain will be. The higher the prior precision, the more strongly the resulting prediction constrains subsequent processing. The relationship also runs in the other direction. How often an individual deploys a given prior over the long term, the cumulative effect of the tendency to deploy priors discussed in Section 3.1.3, feeds back on the precision of that prior. Priors that are repeatedly used and repeatedly confirmed gain precision, whereas priors that go unused or repeatedly fail to predict lose it [14]. Prediction generation and precision weighting are therefore not linked in one direction only, with precision determining prediction. They shape each other.

Interaction 2: a positive cycle between error signal quality and the efficiency of model updating. As Section 3.2.1 showed, an individual with efficient error encoding detects the conflict between literal meaning and context quickly. The resulting error signal is of higher quality, and so is more likely to trigger a model update

© 2026 The Author(s). Authors retain copyright in their articles.

(3.3). The cycle also runs the other way. A model that is updated frequently produces more accurate predictions, which makes the error signals it generates more informative. A deviation from an accurate prediction carries real information, whereas a deviation from an inaccurate prediction tells us very little, because almost any outcome would have been equally unexpected [15]. Efficient error encoding and flexible model updating may therefore reinforce each other in a virtuous cycle.

Interaction 3: the output of model updating becomes a new prior, and so becomes the starting point for the next prediction. This is the most fundamental cycle in the predictive coding framework. Every completed update is, in effect, a revision of the parameters of the internal generative model, and the revised model is where prediction generation begins the next time pragmatic reasoning is required [3]. Model updating therefore matters not only within a single episode of processing. Its cumulative effect continually reshapes the individual's store of prior knowledge. This is the structural reshaping of the generative model by long-term experience discussed at the end of Section 3.3. Once an update has taken place, the basis on which the individual generates predictions in the next similar situation is no longer what it was before.

Taken together, these three interactions constitute a complete predictive loop. Its core logic can be summarised as follows: prior precision triggers prediction generation; failed prediction produces an error signal; error encoding may trigger a model update; the updated model reshapes the priors; and a new round of prediction begins.

4.2 Covariation and Dissociation: A Unified Account of the Three-Component Structure of Pragmatic Inferencing Ability

Drawing on the dynamic loop described in Section 4.1, this section explains, from a predictive coding perspective, why the three components identified by Floyd and colleagues [2] are both correlated and separable. Covariation: shared predictive coding resources. Whether an individual is drawing on social priors, as in social conventions, indirect requests and face management, prosodic priors, as in intonation interpretation, or world knowledge priors, as in causal reasoning, pragmatic reasoning depends on a common predictive coding infrastructure. That infrastructure includes the neural capacity to generate top-down predictions (3.1), the neural efficiency with which prediction errors are detected and encoded (3.2.1), the decision mechanism that dynamically weights signal precision (3.2.2), and the learning rate at which internal models are updated in response to error feedback (3.3). Differences in the parameters of this infrastructure, such as how efficiently errors are encoded, how flexibly precision is weighted, and how quickly models are updated, affect performance across pragmatic reasoning tasks to varying degrees. Some degree of covariation among the three components is therefore an almost unavoidable consequence of how predictive coding works. This is the cognitive source of their covariation.

The sources of dissociation: domain-specific prior content and level-specific precision parameters. If pragmatic reasoning depended entirely on the shared predictive coding infrastructure, the three components should correlate so highly that they would collapse into a single factor. In Floyd and colleagues' [2] data, however, the three components are correlated yet remain independent. This suggests that additional mechanisms must be responsible for their separation. From a predictive coding perspective, those mechanisms operate at two levels.

First, domain specificity of prior content, the distinction drawn in Sections 3.1.1 and 3.1.2. The social conventions component draws on priors about the conventions of social interaction, for example how people use indirect expressions in order to save face [16]. The intonation component draws on priors about the relation between acoustic form and meaning, for example that a rising tone may signal a question and a falling tone an assertion. The causal reasoning component draws on priors about physical and social causal structure, for example that event A usually leads to event B. The prior knowledge underlying the three components barely overlaps in content. Someone with rich social priors need not have finely tuned acoustic priors, and so may be very good at reading between the lines of what a manager says while still being unable to tell from a shift in tone how their partner is feeling. The reverse is equally true. Domain specificity of prior content is the first source of the separation among the three components.

Second, differences in the dominant processing stage and the level of precision. Each of the three components relies on a different stage of the predictive coding cycle, and on precision at a different level. For the social conventions component, the bottleneck lies mainly in the generation and precision of high-level

© 2026 The Author(s). Authors retain copyright in their articles.

priors. If a person's knowledge of conventions such as face management is not finely developed, reasoning will tend to break down even when error encoding and model updating are intact. For the intonation component, the bottleneck lies mainly in the precision weighting of low-level sensory input, the central claim of Section 3.2.2. If a person cannot assign appropriate precision weights to acoustic signals, then even intact high-level social priors will not help. The pragmatic information carried by intonation cannot be passed upward as an informative error signal. For the causal reasoning component, the bottleneck lies somewhere between the two. It depends more on the richness and structural detail of the causal knowledge network in the generative model, the central claim of Section 3.1.2. The three components differ in which stage of prediction is most likely to break down, and this is the second source of their separation.

In sum, the pattern of covariation and separation that Floyd and colleagues [2] found among the three components can be given a unified explanation within the predictive coding framework. Covariation arises from the shared predictive coding infrastructure, and separation arises from domain-specific prior content and stage-specific processing bottlenecks. This integrative framework, of course, still awaits direct empirical test.

5. Future Research Outlook

The preceding chapters set out three cognitive sources of individual differences in pragmatic reasoning from a predictive coding perspective (Chapter 3), and proposed an integrative framework that explains why the three components identified by Floyd and colleagues [2] are both correlated and separable (Chapter 4). The framework proposed here, however, is essentially a theoretical integration and a set of derived hypotheses. Its validity awaits systematic empirical testing. This chapter sets out several testable empirical hypotheses and outlines concrete routes for future research, from behavioural experiments to clinical applications.

5.1 Exploratory Research Directions

Within the theoretical framework proposed here, four directions deserve priority in future research.

The first concerns the relation between social prior precision and pragmatic reasoning. If individual differences in the social conventions component are related to differences in the precision of high-level social priors, future studies could manipulate information about the speaker's social identity, for example by telling participants in advance that the speaker is their superior, their peer or their subordinate, or that the speaker scores high or low on a social anxiety scale. The question is whether this manipulation selectively affects performance on social convention tasks, rather than pragmatic tasks in general.

The second concerns the relation between sensory precision weighting and intonation interpretation. If individual differences in the intonation component are related to differences in the precision weighting of low-level sensory input, then performance on a low-level auditory discrimination task should predict performance on intonation tasks. Does that relationship still hold once general cognitive ability is controlled for?

The third concerns the relation between the efficiency of prediction error encoding and real-time pragmatic performance. If individual differences in error encoding efficiency affect how people perform pragmatically in real time, then performance on language prediction tasks such as Cloze probability judgement should predict how quickly they understand irony and indirect requests.

The fourth concerns the relation between model updating rate and pragmatic flexibility. If individual differences in pragmatic flexibility are related to the rate at which predictive models are updated, then manipulating feedback information experimentally should reveal differences in updating rate between individuals with different working memory capacities. And should those differences predict how quickly a person adapts to a new speaker's pragmatic style?

These directions are not experimental designs but testable hypotheses that follow from the framework. The logic common to all of them is this: if individual differences in the three stages of predictive coding really are the cognitive source of individual differences in the three components of pragmatic reasoning, then these relationships should be observable in future empirical work.

© 2026 The Author(s). Authors retain copyright in their articles.

5.2 Behavioral Experiment Path

These hypotheses can be tested using established individual differences paradigms. One option is to follow the large-scale design of Floyd and colleagues [2], in which the same participants complete three sets of tasks: first, behavioural tasks that tap predictive coding, such as Cloze probability judgement, auditory pitch discrimination, and contextual prediction updating; second, multi-component measures of pragmatic reasoning, covering indirect requests, irony, conversational implicatures, emotional prosody, contrastive prosody and causal reasoning; and third, control measures, including IQ, working memory capacity and theory of mind. Structural equation modelling or path analysis can then be used to test whether the predictive paths from the behavioural indices of predictive coding to the three pragmatic components are component-specific.

5.3 Prospects of Clinical and Intervention Research

The theoretical framework has potential value for the assessment and treatment of pragmatic difficulties in clinical populations.

At the level of assessment, predictive coding offers a new conceptual tool for distinguishing between different types of pragmatic difficulty. Two individuals with ASD who perform similarly on pragmatic tasks may have very different cognitive difficulties. One may mainly lack social priors (3.1.1), while the other may mainly assign excessive precision to sensory input, as in the HIPPEA hypothesis (3.2.2). If behavioural indices, such as the degree of prior reliance shown on Cloze tasks or the sensory sensitivity shown on pitch discrimination tasks, can be used to tell these two groups apart at the assessment stage, it may become possible to target interventions more precisely [7].

At the level of intervention, this distinction maps onto different training strategies.

For missing priors: social stories, scenario simulation and video modelling can be used to enrich the individual's store of social prior knowledge, so that more accurate initial predictions are generated when similar situations arise [1].

For abnormal precision: auditory training or cognitive reappraisal training can help individuals adjust the precision they assign to different signal sources. In irony comprehension training, for instance, participants can be explicitly instructed that when literal meaning and context conflict, they should trust their inference about the speaker's intent rather than the literal wording. This amounts to training the allocation of precision between literal sensory input and social priors [7].

For slow model updating: clear and immediate pragmatic feedback can speed up the rate at which predictive models are updated. Through repeated exposure to similar pragmatic situations, the gains from each update can be consolidated into stable long-term prior knowledge (3.3). A child may at first take "You're so clean!" as a compliment. If the mother immediately frowns, raises her voice, speaks faster and says "Go and have a bath", and if an adult then points out that when literal meaning conflicts with the situation and is accompanied by a particular prosody and particular actions, the utterance should be read as irony, the child's predictive model can be updated. After several such experiences, the child internalises the rule as a stable prior and comes to recognise irony quickly.

6. Conclusion

This paper has examined the cognitive origins of individual differences in pragmatic reasoning from a predictive coding perspective. In response to the three-component structure of pragmatic language use identified by Floyd and colleagues [2], namely social conventions, intonation and world knowledge–based causal reasoning, this paper has advanced and defended a core hypothesis: individual differences in predictive coding are a common cognitive source of the individual differences found across the three components of pragmatic reasoning.

In developing this hypothesis, the paper first set out the core mechanisms of predictive coding and decomposed them into three key stages: prediction generation, prediction error encoding and precision weighting, and adaptive updating of predictive models. It then argued, stage by stage, how differences between individuals at each of these stages shape the overall pattern of individual differences in pragmatic reasoning.

© 2026 The Author(s). Authors retain copyright in their articles.

At the prediction generation stage, differences in the richness and precision of social priors, in the structure of world knowledge priors, and in the pragmatic sensitivity with which priors are deployed determine where pragmatic reasoning starts and in which direction it proceeds.

At the prediction error encoding and precision weighting stage, differences in the neural efficiency of error encoding and in the flexibility of precision weighting determine whether a person can effectively detect the conflict signal that arises when literal meaning and context do not match, and can assign it an appropriate processing weight.

At the adaptive model updating stage, differences in updating speed, updating depth and the shaping effects of long-term experience determine how flexible and adaptive a person's pragmatic reasoning is.

The paper then proposed an integrative framework. The three stages, it argued, do not operate in isolation. They form a dynamically interacting predictive loop: prior precision initiates prediction generation, prediction failure produces an error signal, error encoding triggers model updating, the updated model reshapes the priors, and a new round of prediction begins. This framework offers a unified explanation of the pattern of covariation and separation among the three components. Covariation arises from the shared predictive coding infrastructure, and separation arises from domain-specific prior content and stage-specific processing bottlenecks.

Finally, the paper relocated existing explanatory perspectives, including theory of mind, language experience and executive function, within the predictive coding framework. Each of these perspectives corresponds to a particular stage or parameter of predictive coding, and predictive coding in turn offers a possible underlying framework in which they can be integrated. This attempt at theoretical integration offers a new explanatory framework for the study of individual differences in pragmatic reasoning, and it points future empirical work towards questions that can actually be tested.

As a theoretical review, of course, the core hypothesis advanced here still awaits systematic empirical testing. Whether predictive coding can serve as the underlying framework that unifies these perspectives depends on whether it gains consistent support across behavioural experiments, neuroimaging and clinical intervention. It is hoped that the framework proposed here will prove useful to future research, and will help move the study of individual differences in pragmatic reasoning from listing factors towards integrating mechanisms.

References

[1] Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development, 14*(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

[2] Floyd, S., Jouravlev, O., Poliak, M., Mineroff, Z., Gibson, E., & Fedorenko, E. (2025). Three distinct components of pragmatic language use: Social conventions, intonation, and world knowledge–based causal reasoning. *Proceedings of the National Academy of Sciences of the United States of America, 122*(50), Article e2424400122. https://doi.org/10.1073/pnas.2424400122

[3] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

[4] Zhang, X., & Zhang, D. (2025). Neural computation and modeling of predictive coding in natural language comprehension. *Journal of Psychological Science, 48*(4), 861–875. https://doi.org/10.16719/j.cnki.1671-6981.20250409

[5] Jano, S., Cross, Z. R., Chatburn, A., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2024). Prior context and individual alpha frequency influence predictive processing during language comprehension. *Journal of Cognitive Neuroscience, 36*(9), 1898–1936. https://doi.org/10.1162/jocn_a_02196

[6] Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121*(4), 649–675. https://doi.org/10.1037/a0037665

© 2026 The Author(s). Authors retain copyright in their articles.

[7] Vicente, A., Michel, C., & Petrolini, V. (2024). Literalism in autistic people: A predictive processing proposal. *Review of Philosophy and Psychology, 15*, 1133–1156. https://doi.org/10.1007/s13164-023-00704-x

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

[9] Levinson, S. C. (1983). Conversational structure. In *Pragmatics* (pp. 284–370). Cambridge University Press.

[10] Marocchini, E., & Domaneschi, F. (2022). "Can you read my mind?" Conventionalized indirect requests and theory of mind abilities. *Journal of Pragmatics, 193*, 201–221. https://doi.org/10.1016/j.pragma.2022.03.011

[11] Esteve-Gibert, N., Schafer, A. J., Hemforth, B., Portes, C., Pozniak, C., & D'Imperio, M. (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory & Cognition, 48*, 566–580. https://doi.org/10.3758/s13421-019-00990-w

[12] Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

[13] Bornkessel-Schlesewsky, I., Sharrad, I., Howlett, C. A., Alday, P. M., Corcoran, A. W., Bellan, V., Wilkinson, E., Kliegl, R., Lewis, R. L., Small, S. L., & Schlesewsky, M. (2022). Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating. *Frontiers in Psychology, 13*, Article 817516. https://doi.org/10.3389/fpsyg.2022.817516

[14] Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

[15] Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry, 4*, Article 47. https://doi.org/10.3389/fpsyt.2013.00047

[16] Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

© 2026 The Author(s). Authors retain copyright in their articles.