

An Integrative Literature Review Aligning Cross-disciplinary Continua for a Computational *raison d'exprimer*

Antonio Origlia*
Università di Napoli Federico II

Maria Di Maro**
Università di Napoli Federico II

*The landscape of Conversational Artificial Intelligence has become extremely varied, in the last years, as the capability of using language in machines has become the main showcase for generative applications. From a language philosophy point of view, however, the reason why Generative AI produces linguistic content is not aligned with the one motivating humans. In this paper, we provide a multidisciplinary integrative literature review spanning linguistics, cognitive science and computer science to propose a systematic way of organising this knowledge around *raison d'exprimer*: the reason why a machine should use language. We present both a horizontal model describing how different aspects of communication blend into each other, the hypertriangle of communication, and a vertical model providing a common axis to align multiple disciplines involved in Conversational AI: the illocutionary gradient. By showing how concepts belonging to multiple disciplines align themselves in these models, we provide an extensible theoretical tool to study conceptual alignments in different fields.*

1. Introduction

Recent directions in Artificial Intelligence stress the need of accessing knowledge found in different scientific fields, not necessarily technological, to push forward progress in simulating human intelligence processes. The intrinsic cross-disciplinary setting scholars need to move into, therefore, is very complex and it ranges from humanities to mathematical domains. In this paper, we propose a way to organise the available knowledge to support comparisons among different disciplines and topics while avoiding strict classification systems. We invest, instead, in stressing the role of *continuous* representations both in terms of different moments involved in communication blending into each other, a horizontal view, and in terms of different levels of reasoning enabling communicative strategies, a vertical view. This is important to overcome the current impasse depending on machines generating linguistic content without an communicative goal. In fact, although Artificial Intelligence has made remarkable progress, what appears particularly compelling, against the backdrop of growing interest in Conversational AI and its longstanding role in the attempt to emulate human thought processes, is the notion of *raison d'exprimer*.

Linguistics considers the internal motivation behind speaking as *illocutionary force*. However, Linguistic studies can only focus on analysing the external manifestations of it: a Conversational AI perspective allows to explore the problem by simulating this

* E-mail: antonio.origlia@unina.it

** E-mail: maria.dimaro2@unina.it

driver in an accessible system. While we cannot claim that the way *raison d'exprimer* corresponds to how this, eventually, exists in humans, we provide machines with internally modelled motives underlying language production as a behaviour, grounded in linguistic theory. This also offers a perspective that may help us move beyond the superficial illusions generated by technological artifacts because of the anthropomorphisation bias. In fact, the reactive nature of Generative AI raises questions over the actual capability of these systems to go beyond speaking without intending to say anything. First, let's highlight three main aspects of generative models used for dialogue management:

- they only capture surface (locutive) aspects of communication because they are trained to predict the most probable continuation of some textual input;
- their only intention or goal, if any, is to complete the text: they don't autonomously pursue communicative goals beyond text completion, thus exhibiting a limited form of illocutionary motivation;
- they operate reactively: being connectionist models, no matter how complex, they are not conceptually different from Braitenberg vehicles (Braitenberg 1986), a classic example to describe attribution of intelligence, because of the anthropomorphisation bias, to wired behaviours.

The last comparison concerns not complexity but the epistemics of attribution. Braitenberg's vehicles make a methodological point, captured in his law of uphill analysis and downhill synthesis: behaviour that an observer spontaneously describes in psychological terms, fear, aggression, foresight, can be produced by mechanisms that contain none of the states so attributed, and an observer reasoning from behaviour back to mechanism will systematically over-attribute interior complexity. The vehicles are a demonstration that the appearance of mentality is, in large part, an observer-side effect. While the difference in complexity is extreme, our note is that if the illusion arises even for systems of trivial internal organisation, it arises more readily as the apparent complexity of the system grows. The complexity of LLMs actually sharpens the warning the parallel is meant to convey. We refer to Braitenberg not to suggest that these systems are simple, but to observe that the human disposition to read mentality into behaviour is powerful enough to be triggered by the barest of mechanisms, and is correspondingly more potent in the presence of fluent linguistic output.

On these bases, Generative AI is a powerful tool for Natural Language Generation, and in Conversational AI it should be treated as such. The degree to which genuine dialogue emerges tracks the degree to which communicative acts are connected to a system's own goals, which is a foundational point for a significant amount of cross-disciplinary research involving linguistics and AI in the early '90s, although usually interpreted in a categorical way: we will reframe this position in a gradualist way in this paper. A primary motivation of this stance is the seminal work of Austin, highlighting that "To say something is to do something; or which by saying something or in saying something we are doing something.". This is notably close to Judea Pearl's view towards the relevance of *doing* in AI with respect to mere *observing*, in the *ladder of causation* and supports a view that speech acts must be modelled as *actions* rather than machine learning *reactions*. This foundational concept will be discussed more in detail throughout the paper and represents a basic organisation principle for the models we propose.

To introduce account for intentionality in Conversational AI and look beyond the apparent presence of judgement (Loru et al. 2025) in its processes, it is necessary to go

back to the performative nature of speaking. That is, to explicitly consider, once again, speech acts as means to pursue ends and reach goals, as opposed to mere reactions to input stimuli. This implies the necessity to “[...] recognize the multi-dimensional nature of language use, ranging from cognitive to emotional facets, from actional to affective ones, from the social to the subjective. The aim is to account for language use in all its irreducible complexity” (Sbisà 2001, p.1792). A truly performative view of language is, by its very nature, cross-disciplinary.

Specifically, we propose a model of spoken communication inspired by linguistic theory with recognisable cognitive abilities that can be implemented with computational tools. These models are meant to provide an orienting framework for scholars in the complex multidisciplinary fields ranging from humanities to psychology and computer science. Terminological issues and overlapping information can be difficult to disentangle and yet it is now imperative to conduct this kind of research to avoid become victims of the statistical illusions we are now able to create. Moreover, we provide a reference alignment dimension, the illocutionary gradient, around which the view we present here can be extended in the future to other aspects not considered in this work. While we limit our discussion to illocutionary strength, here, we hope that it will be possible to generalise this approach up to the level of intentionality for AI.

The paper is organised as follows: Section 2 describes the literature review methodology we adopted to organise the relevant works, Section 3 presents the hypertriangle of communication, supporting the placement of cross-disciplinary concepts in different moments of the communicative exchange, highlighting adjacencies as a way to describe how these blend into each other. Cross-disciplinary concepts of interest are, then, presented in Section 4 as part of multiple continua and aligned to a common dimension, the illocutionary gradient. On each continuum, discrete reference concepts, called *anchors*, are identified to support a comparative analysis of concepts found on different continua and aligned through the gradient to verify their consistency with each other. For each continuum, following the order of presentation, a dedicated Comparison Table is presented in the Appendix to discuss the relationship of *anchors* on that continuum with the previously described ones. This way, comparisons are discussed systematically in a systematic way. In Section 5, the considered continua are compared with a hierarchy of capabilities in a Conversational AI architecture to verify the alignment of the considered theories with technological modules.

2. Methodology

To propose a theoretical model supporting scholars in comparing theories coming from different fields, we conducted an integrative literature review. The goal of this process is to “[...] critically analyse and examine the literature and the main ideas and relationships of an issue” (Snyder 2019, p.336). A key difference with other kinds of literature review is that integrative ones “[...] should not be descriptive or historical but should preferably generate a new conceptual framework or theory”. A summary of the characteristics of different kinds of literature review are reported in Table 1, as described in Snyder (2019).

The goal of this effort is to provide a clear framework to organise concepts and relationships involved in Conversational AI from multiple perspectives and how they may relate to each other. This may help researchers with different backgrounds compare theoretical principles involved in the process of speaking with intention to support a more coordinated discussion around the topic. This will hopefully help develop explainable model-based approaches to Conversational AI that also take advantage

Table 1
Approaches to literature review

	Systematic	Semi-systematic	Integrative
Typical purpose	Synthesize and compare evidence	Overview research area Track development over time	Critique Synthesize
Research questions	Specific	Broad	Narrow or broad
Search strategy	Systematic	May or may not be systematic	Usually not systematic
Sample characteristics	Quantitative articles	Research articles	Research articles Books Other published texts
Analysis Evaluation	Quantitative	Qualitative Quantitative	Qualitative
Examples of contribution	Evidence of effect Inform policy and practice	State of knowledge Themes in literature Historical overview Research agenda Theoretical model	Taxonomy Classification Theoretical model Framework

from modern machine learning tools. This is consistent with the goal of an integrative literature review to synthesize the literature, intended as “[...] a creative activity that produces a new model, conceptual framework, or other unique conception” (Torraco 2005, p.362).

The guiding principles to structure this work were cross-referencing and continuation. Cross-referencing means that we will strongly highlight relationships between analysed concepts, rather than concepts themselves. Continuation means that we will concentrate on building models that do not pose strict separations among their components and will make a strong use of *continua* and blends. Accordingly, distinctions discussed throughout the paper should be interpreted as relative differences in complexity and functional organisation rather than as absolute ontological boundaries. We will discuss relationships both on a peer level and on a hierarchical level to provide a more encompassing view. The proposed theoretical framework is composed of two models:

- The hypertriangle of communication: a horizontal model of the relationships among the main elements involved in spoken communication;
- The illocutionary gradient: a vertical model aligning six different *continua* from psychology, linguistics and computer science using illocutionary force as a shared dimension.

Together, they aim to reveal cross-disciplinary regularities and offer a unifying framework for concepts often treated in isolation.

As the goal is to describe a model for a computational *raison d'exprimer*, in selecting the literature material we included in our analysis we concentrated on theories that either come from the computer science domain or have found applications in computational settings. Preference has been given to theories that support, either natively or implicitly, a hierarchical view of processes involved in communication. This is not a strong assumption for a computationally oriented approach, as cognitively inspired architectures are often organised around hierarchical principles. The general rule, in these models, is that lower processes cannot perform tasks that pertain higher ones while these build upon the information provided by lower ones.

Concerning the horizontal view, we provide a visual representation of the relationships between fundamental macro-elements related to communication and use this structure to motivate the cross-disciplinary framework we propose in this paper. This representation is obtained by extending a classic approach to the representation of linguistic symbols to the objects they represent: the semiotic triangle (Ogden and Richards 1927). We start from this model because, from a computational modelling point of view, it can be formally represented as a graph. Graphs are an advantageous approach to the representation of computational models because of their flexibility in representing complex networks of relationships while also being human readable. However, since the semiotic triangle is a *closed* model, limited to a narrow view of semantic interpretation, we extend it through the use of **adjacent** triangle models to address these problems. Through adjacency, we avoid creating categorically separated models covering different aspects of the process of dealing with the environment through linguistic actions and highlighting continuity between processes and how they blend into each other.

Concerning the vertical view, our coverage focuses on three main areas: linguistics, psychology and computer science. The approach aims at using a common dimension, the illocutionary gradient, to support comparisons between concepts belonging to different disciplines, using the reason why we communicate to align theories describing how we do it. The continua we discuss here are not meant to be exhaustive: the proposed model allows extensions and future comparisons through the use of the shared dimension. A graphical representation of these concepts is presented in Figure 1. Continua represent the different perspectives that can be considered when dealing with the emergence of communicative abilities. On each continuum, we represent the *relative* complexity degree of abilities and representations used in communicative behaviour. Of course, there are many other continua that may be taken into account for such an effort. While we do not claim that the current model is exhaustive, we believe we covered the dimensions where overlaps in the literature are stronger so that the usefulness of our work in cross-disciplinary applications from the community is high. We also consider the illocutionary gradient a powerful reference dimension to align more continua in the future, if necessary. The continua we selected are:

- Sensibility (psychology): represents the degree of complexity of sensible experiences, from physical (dis-)pleasure to moral judgments;
- Drivenness (psychology): reflects the extent to which communicative acts are aimed at achieving various types of goals. It differentiates between goal-oriented and goal-directed behaviour, while also taking into account behaviour that adapts to external goals;

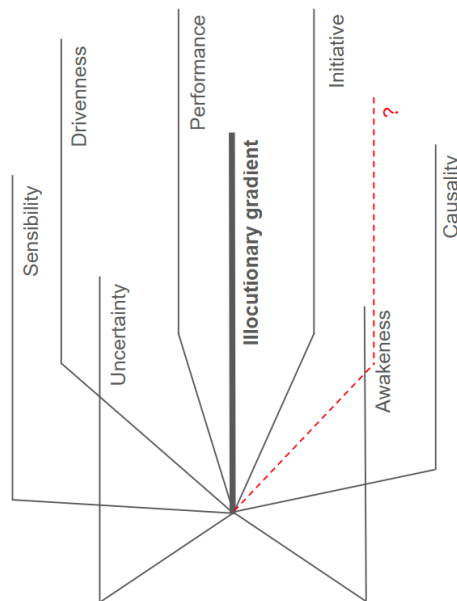


Figure 1

Through the common dimension of the illocutionary gradient, multiple theories from different disciplines may be compared in a critical analysis. Furthermore, other aspects not considered in this paper may be added in the future, in the same way. This supports easier comparisons between disciplines as, by aligning a new continuum to the illocutionary gradient, analyses to be performed with other disciplines are automatically obtained.

- Causality (computer science): represents the degree at which causal relationships are understood and exploited;
- Uncertainty (computer science): represents the degree of predictability in processes and data contributing in producing the communicative act;
- Initiative (linguistics): represents the type of intervention a system applies to the given context, reflecting the different ways a speaker may act upon it to achieve a desired configuration. From a linguistic perspective, this corresponds to the performative use of speech acts aimed at pursuing specific goals;
- Performance (linguistics): represents the degree at which communicative acts imply the disclosure of the speaker's internal state or access to the other's.

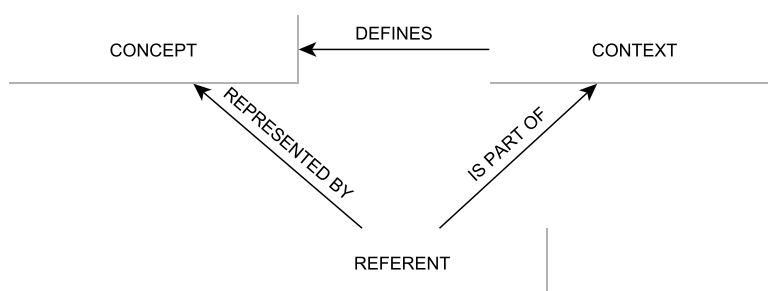
For each considered area, two *continua* have been selected to foster comparisons and critically analyse the relationships arising among concepts after aligning them to the common dimension. After aligning each continuum to the illocutionary gradient, we discuss the consequent relationships with other continua as part of our critical analysis. By doing so for six, relevant topics for the case of Conversational AI, we demonstrate that the illocutionary gradient can serve as a reference for scholars to synchronise their efforts and position their contributions with respect to the others.

3. The hypertriangle of communication

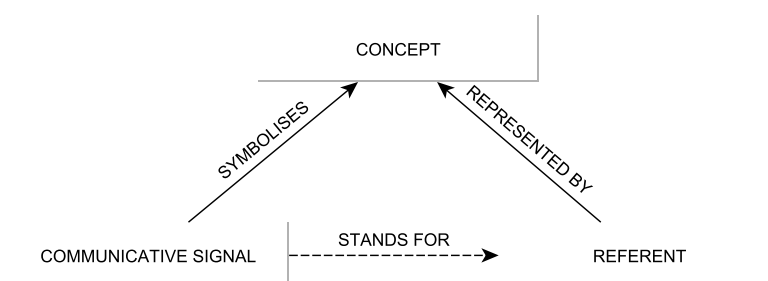
While useful as starting point for a computationally interpretable model, the semiotic triangle suffers by relevant limitations due to its restricted focus on a tripartite relation between *symbol*, *referent*, and *thought or reference*. While this captures an important dimension of meaning-making, it excludes the dynamic, relational, and systemic aspects of language use. As Saussure's theory of the linguistic sign highlights, meaning is not inherent in the link between a symbol and its referent but emerges from the differential relations between signs within the system of language (de Saussure 1916). Moreover, in his "circuit of speech" (*circuit de la parole*), Saussure described communication as a cyclical process involving both interlocutors: the speaker's production of an acoustic image (phonation), its transmission as a sound wave, and the hearer's perception and interpretation of it, before the cycle restarts in the reverse direction. This dynamic conception stresses the role of interlocutors and the continuous flow of communicative acts, absent from the semiotic triangle's static representation. Hjelmslev's glossematics provides yet another extension of this perspective. In his framework, language is analysed in terms of two interdependent planes: the *expression plane* (forms of signifiers) and the *content plane* (forms of signifieds), each of which can be further divided into form and substance (Hjelmslev and Whitfield 2024). This layered model underlines that meaning does not reside only in reference to external objects but also in the internal structural relations of the semiotic system itself. Importantly, however, these internal relations are not purely abstract: they are always mobilised in processes such as experience, reasoning, and interpretation. In other words, the structuring of expression and content is constantly re-actualised in communicative practice, where interlocutors map systemic relations onto lived experience and cognitive activity. Recognizing this processual dimension allows us to connect Hjelmslev's structural focus with the dynamics of meaning-making in actual communicative contexts. Incorporating these perspectives allows for a more comprehensive model of communication: one that is not only graph-based and referential but also relational, systemic, and processual. In this sense, the extension of the semiotic triangle into a *hypertriangle of communication* is aligned with the structuralist insight that meaning arises through networks of relations, thereby offering a richer framework for modelling communicative interactions.

On these premises, we propose a set of six adjacent triangles dealing with different aspects of communication, forming the *hypertriangle of communication*. From a strictly geometric point of view, the juxtaposition of six triangles forms an hexagon but, since a hexagon is not usually represented with the internal vertex and edges needed for our representation, we opted for another term. This is meant to stress how components shared among multiple triangles are related across different dimensions of a communicative exchange. This approach aims at going beyond the boundaries of the semiotic triangle in the sense that it extends the interpretation of the meaning in the communicative setting.

The first triangle is the *framing* triangle, shown in Figure 2. The first goal of our model is, in fact, to stress the importance of context, and therefore pragmatics, in representing the communication exchange. Through framing, the *context*, defines the *concept* among the different ones that the *referent* can be represented by. *Context* is defined as the set of conditions in which the communicative exchange takes place (Catford 1965) while the *concept* is a mental image or generalisation existing in the mind of a speaker and the referent is the object in the physical reality (Ogden and Richards 1927). This

**Figure 2**

The framing triangle, stressing the importance of pragmatics in our representation of communication concepts and their relationships.

**Figure 3**

The referring triangle, directly related to the semiotic triangle.

accounts for the possibility of the same referent being conceptualised in different ways in different contexts¹.

The *referring triangle*, shown in Figure 3, has the same role as the semiotic triangle but contains a *communicative signal*, a physical behaviour that can be detected by sensory apparatuses, instead of the original *sign*. In fact, the semiotic triangle was not meant to represent the details of the perception and interpretation process leading to the detection of the *sign*, nor its internal use to shape a response. In our case, instead, we wish to provide a more detailed representation of the macro-elements related to the interpretation of communicative signals. The *communicative signal* symbolises a *concept*, possibly representing the *referent* if this has a physical nature². The dashed line, in fact, refers to the indirect and arbitrary relationship between the sign (communicative signal in our model) and the referent (de Saussure 1916).

The *perceiving triangle*, shown in Figure 4, stresses the importance of *perception* during communication. We define *perception* as the ability to detect and process signals at different granularity in real time. Explicitly considering *perception* is important because

1 Example: a toilet in an art exhibit or in a bathroom is still a toilet, but in one case it is conceptualised as an art installation and in the second one as relief device.

2 Example: The sounds composing the word *apple* symbolise the concept of the actual fruit by standing for it during communication.

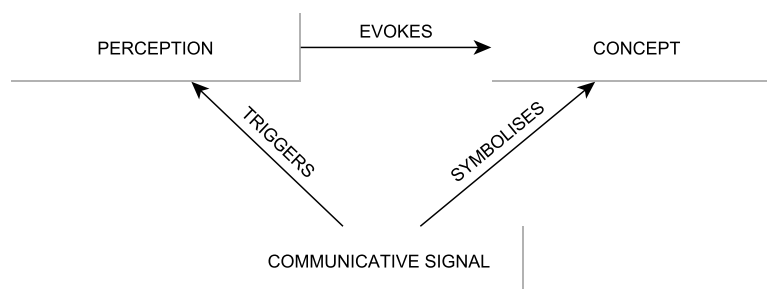


Figure 4

The perceiving triangle, stressing the importance of perceptual processes in the communicative exchange.

early processes involved in communication filter and enhance specific aspects of the *communicative signal* to support internal processing, thus influencing this. Gestalt theory (Wertheimer 1938), ecological theory (Gibson 2014) for vision, and rhythm-based speech scanning (Poeppel and Assaneo 2020) are some examples of theoretical frameworks dealing with perception processes that are closely related to the formation of meaning. The communicative signal used to symbolise a *concept*, therefore, triggers *perception* in the listener, evoking *concepts* consistent with the perceived *signal*³.

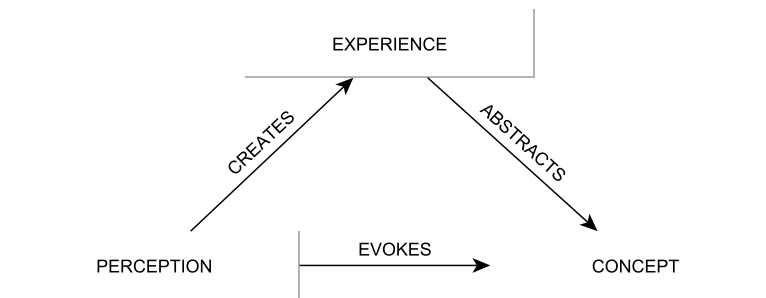
The *representing triangle*, shown in Figure 5, focuses on the role of episodic memory (Tulving 1972) in abstracting and evoking *concepts* using past *experience*, the collection of percepts and/or living through facts and events. *Perception* creates *experience*, possibly using self-monitoring to establish whether a new episode is worth remembering. Emotional memories and events disclosing causal relationships are examples of *memorable* episodes. By evoking abstracted *concepts*, relevant information is retrieved associatively in the network of past episodes, integrating the *concept* with its correlates. This can be considered a form of semantic (pre-)activation (Federmeier and Kutas 1999)⁴.

The *reasoning triangle*, shown in Figure 6, concentrates on the way evoked *concepts* and past *experience* with them get a role in shaping the generated *communicative act*. The role of memory is “not recall but preparation” (Huron 2008, p.219), so past *experience* may influence how the *communicative act* will be shaped and which concepts will be used to pursue its goal. Since this relationships is not always active, we indicate it using a dashed line. Since the process of remembering also retrieves concepts that are semantically related to the ones directly evoked by perception, this triangle represents how concepts that were not previously mentioned, but are related to the situation, may be used to support the preparation of the response to form the next dialogue move⁵.

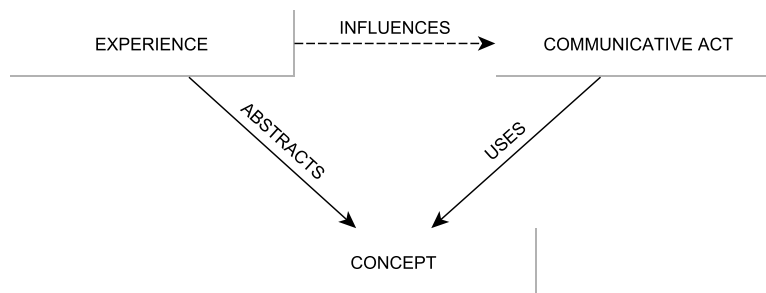
³ Someone says the word *apple*, someone else hears it, processes it and imagines the fruit or the IT company.

⁴ Example: Someone often perceives the word *engineer* associated with male figures, so when the word is perceived again it will evoke the concept of someone doing technical work integrated with the characteristic of being male. Gender bias may be a side effect of the interaction between the actors involved in this triangle.

⁵ Example: *Cut the butter* coming after *Melt the butter* generates a clarifying intention leading the listener to say *Didn't I have to melt the butter?* (Di Maro, Origlia, and Cutugno 2021a). While the butter is referenced, in the sentence, it goes against widely shared past experience (Common Sense), to specify the concept that liquids cannot be cut. So, in choosing which concepts to use, the speaker chooses to omit the reason why the action cannot be performed because this can be reconstructed by the interlocutor.

**Figure 5**

The representing triangle, describing how memorised past experience may integrate currently evoked concepts with relevant associated information created by past episodes.

**Figure 6**

The reasoning triangle, showing how the evoked concepts and past experience with them influences how the next communicative act is shaped.

The *acting* triangle, shown in Figure 7, highlights the performative nature of communicative acts. It shows how using concepts in the specific context where the communicative exchange is happening alters this in a goal-directed way. The pursue of goals through the generation of communicative acts makes these subject to intentions, which are stable in the sense that they are conscious and always consistent with each other (Searle 2003). Also, intentions are strongly linked to behaviour (Ajzen and Fishbein 1980) “[...] in that they imply a decision of pursuit” (Miceli and Castelfranchi 2014, p.6). Their link in this triangle specifies how the existence of communicative acts serves purposes and therefore needs a clear representation of *why* they are generated. As we will argue in Section 5, reasons to communicate may be of very different complexity and even more or less intentional, but, in general, they are always there ⁶.

The juxtaposition of the six triangles we described creates the high-level representation of the relationship network linking the main macro-elements forming the hypertriangle of communication, shown in Figure 8. Through adjacent triangle edges, this representation shows the relationships between macro-elements involved in com-

⁶ Example: a person wants to convince the interlocutor to watch a movie and highlights the presence of a certain actor, known to be liked by the interlocutor, to maximise the probability that the proposal will be accepted.

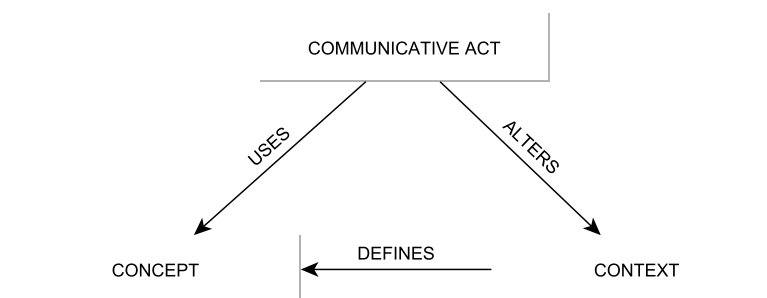


Figure 7

The acting triangle, highlighting how the capability to use concepts in the communicative context with the specific goal of altering this provides a reason why communicative acts exist.

municative exchanges and their roles during different moments. The hypertriangle, as a composition of triangles, does not advocate for strong discretisation between different moments involved in the communicative exchange. On the contrary, adjacencies represent blended boundaries between these moments, as opposed to, for example, a block diagram.

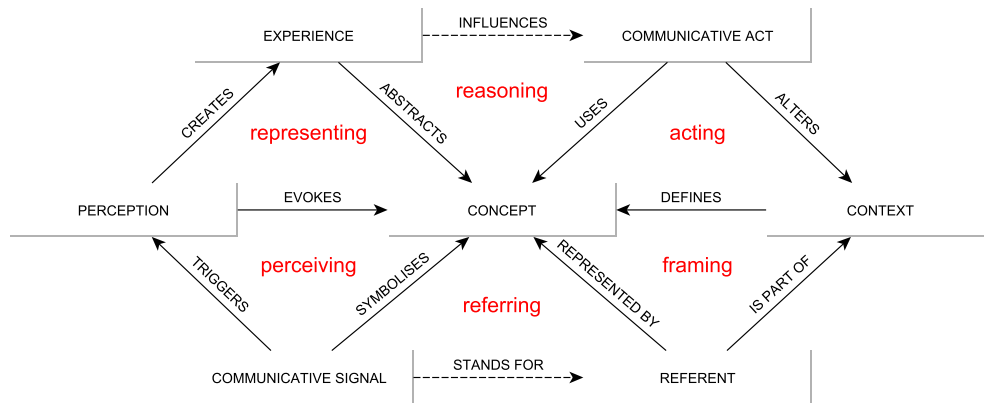
Moreover, while it is tempting to identify an ordered sequence of moments starting from framing and going to acting counter-clockwise, we advise not to interpret the representation like this, as it has been well studied, now, that the process of continuously perceiving communicative signals and the surrounding context may lead to incremental models of meaning that are subject to revision as information arrives. Also, not all processes included in the proposed triangles are all always necessarily involved in the perception/action cycle. For example, reflexes and conditioned behaviours do not need capabilities concerning *representing* and *reasoning*.

The *concept* macro-element is meant to be clearly central to highlight the importance of the making of meaning and symbolic manipulation during communication, while avoiding a clear cut separation with the physical processes linked to perception and action. This is important to consider different reasons why communicative acts may be performed in terms of their illocutionary characteristics, being them the result of high-level, complex reasoning or lower-level, simpler reactions to stimuli.

Lastly, triangles are organised to form dyads between the external context and internal processes, so that *perceiving* and *acting* represent the classic perception/action cycle. The *referring* processes involving the interlocutor correspond to internal *reasoning* over them to produce reactions. The contextual reality of *framing* corresponds to its internal approximation by *representing*.

4. The illocutionary gradient

Theoretical models across various disciplines tend to use discrete classes to describe phenomena of interest. While this allows to identify and discuss the macro-elements of these phenomena, as research goes deeper into them it is often the case that newer models decompose previous classes into their components, compiling dimensional models. Through these, new theories escape the limitations of named classes in representing blends between concepts. This happened, for example with the shift in the representation of emotions from the categorical model (Ekman 1999) to various kinds

**Figure 8**

The juxtaposition of the six triangles produces the hypertriangle of communication.

of dimensional spaces (e.g. (Fontaine et al. 2007)). The case of emotions is particularly exemplar as some recent models highlight the role of ordinal relationships between different experiences on the same dimension, as opposed to absolute measures, in shaping the experience (Yannakakis, Cowie, and Busso 2017). This is consistent, in particular, with the relative judgment model described in Stewart, Chater, and Brown (2006), where decisions are taken on the basis of pairwise comparisons between samples retrieved from memory and the subsequent ranking. Following this approach, we consider multiple continua from different theories and we align them for the case of spoken communication management. In doing this, we use the internal illocutionary strength (Sbisà 2001) as a common ordering dimension: the illocutionary gradient. Each one of these axes involves different phenomena that can be organised using the illocutionary gradient as a common dimension.

We highlight that the lower limit of the gradient is still a value on the scale, not a region beyond it. A claim that a given system has “no raison d’exprimer” should accordingly be read as the claim that it sits at or near this lower bound, in the same way that “off” is a setting of a continuous dimmer rather than the negation of the dimension along which it varies. Several of our claims about contemporary systems, if read as categorical demarcations, would sit uneasily with the gradualist view we support.

In the rest of this Section, we summarise theoretical aspects related to each continuum and describe how they can be aligned to the illocutionary gradient. Selected continua are not meant to be exhaustive and the considered theories have been selected across the disciplines of interest among the ones that describe, either explicitly or implicitly, a hierarchical conceptual organisation. To better describe the nature of the continua, we also select a number of discrete labels, acting as *anchors*, to identify reference concepts we can use for exemplification and cross-disciplinary comparisons. Through *anchors*, we describe how we rank these concepts with respect to the illocutionary gradient. As we present each continuum, we will compare its *anchors* with the ones in the previous continua to exhaustively discuss the alignments emerging through the illocutionary gradient. Comparison Tables in the Appendix present these aspects. Also, we will link the description of the continua to the hypertriangle of communication to establish the theoretical background for the analysis in Section 5, where we will discuss

the relationship of the theoretical models we present here with a Conversational AI architecture.

Each continuum concerns a wide amount of literature and exhaustively summarising each area is beyond the scope of this paper. Our discussion aims at highlighting contact points, aligned using the illocutionary gradient, to support the development of cross-disciplinary approaches to Conversational AI. Theories we consider in our discussion are used to exemplify how cross-disciplinary comparisons can be performed methodically through the shared dimension and are characterised by a hierarchical organisation of concepts. This is not a strong assumption as this usually a solid approach to order cognitive abilities in general. In cognitive science terms, this approach follows closely what dual-process theories describe as the continuum between System 1 and System 2 (Kahneman 2011).

4.1 Sensibility

The most fundamental continuum we consider is related to how a speaker experiences sensations while participating to a dialogue. The role of emotions in shaping intelligence is now well established as a fundamental component of cognition (Miceli and Castelfranchi 2014) and widely considered even in technological applications in the field of Social Signal Processing (Vinciarelli, Pantic, and Bourlard 2009; Burgoon et al. 2017) but, historically, the concept of emotions has been quite elusive to define. In particular, while it was well understood that a difference exists between *core affect* (Russell and Barrett 1999) and more structured experiences, terminological issues have caused difficulties in comparing studies and models. In our case, these distinctions are particularly important because of the need to relate sensible experiences to different degrees of illocutionary strength.

First of all, we highlight that non-conscious sensible experiences do not directly activate conscious processes leading to exhibiting behaviour. Consider, for example, the case of *proto-fear*, as defined in Miceli and Castelfranchi (2014, p.21): “[...] a basic, hardwired reaction of alarming displeasure that does not necessarily involve any cognition”. This kind of sensible experience that does not depend on previous exposure to the eliciting stimulus must be aligned with low illocutionary levels, directly dealing with low-level processes whose *results* may produce automatic preparatory reactions. In this sense, core affect does not directly influence higher levels of cognitive processes related to communication but the result of the unconscious processes it triggers may do so. While sensible experiences that can be related to core affect are primarily concerned with the *perceiving* triangle, pre-conscious and conscious ones are concerned with the *representing* and *reasoning* triangles. In particular, past experiences with incoming stimuli elicit reactions that both inform about the situational evolution and interact with personal goals to shape behaviour. This set of sensible experiences are more closely related to appraisal theories of emotions (Scherer 2001). The difference between these experiences appears to be clear cut so that different models can be applied depending on what is being represented. This is not, however, compatible with our approach based on continua because of the need to avoid establishing categorical boundaries. For this reason, we follow the cognitive-motivational approach from Miceli and Castelfranchi (2014), which accounts for both non-evaluative and evaluative appraisal by considering goals and beliefs as direct components of emotions, rather than a derivation of appraisal. The main difference lies in relaxing the assumption that evaluative beliefs are necessary for emotions that pertains appraisal. In the cognitive-motivational approach, only factual beliefs and goals are considered necessary for emotions to emerge. Moreover, dif-

ferentiating this approach from the belief-desire theory of emotions (Reisenzein 2009b, 2009a), where a causal link between beliefs/desires and emotions is identified, the cognitive-motivational model proposes an integrated view of the phenomenon, so that beliefs and goals are actually components of emotions rather than distinct mental states. A theoretical framework accounting for sensible experiences with different degrees of conscious involvement and proposing a strong integration with beliefs and goals is fully compatible with the objectives of this work.

Given the terminological inconsistencies in the literature related to emotions, in identifying the explanatory anchors we use for this continuum we refer to the survey provided in Munezero et al. (2014). This work, although specifically targeted at text analysis, provides a clear view of sensible experiences that we use as categorical anchors for this continuum. Anchors are ordered by illocutionary strength in the following of this Section.

On the basis of Shouse (2005), *affect* is defined as “[...] a non-conscious experience of intensity: as a moment of unformed and unstructured potential”. Affect reportedly being always related to non-conscious experience (Massumi 2021) is compatible with our low-illocutionary view of this anchor: the lowest among the considered ones. Also, affect is described as “[...] the body’s way of preparing itself for action in a given circumstance by adding a quantitative dimension of intensity to the quality of an experience”, which is compatible with the preparatory role of core-affect in Miceli and Castelfranchi (2014), where goals in its emergence are actually pseudo-goals (see Section 4.2). Again, affect appears to be mostly related to the *perceiving* triangle.

Drawing again from Shouse (2005), *feelings* are “[...] a sensation that has been checked against previous experiences and labelled. It is personal and biographical because every person has a distinct set of previous sensations from which to draw when interpreting and labelling their feelings”. Feelings are, therefore, conscious experiences that mainly relate to the *representing* triangle in that they interact with memory in creating and retrieving experiences with the observed situation on a purely reactive basis.

Emotions are defined in Batson, Shaw, and Oleson (1992) as “[...] reflections of the existence of a specific goal or of perceived change in one’s relationship to a specific goal in the present”. This view of emotions related to goals and perspective changes towards them is compatible with their representation in the cognitive-motivational model. Furthermore, Kleinginna and Kleinginna (1981, p.355) defines them as “[...] a complex set of interactions among subjective and objective factors, mediated by neural and hormonal systems”. This aspect highlights that emotions go beyond simple perception and, while considering subjective experience with the eliciting stimulus, they are the result of the joint evaluation of both subjective and objective experience. Thus, they influence the basic processes involved with the *reasoning* triangle. In a discussion aimed at clarifying differences between sensible experiences, Scherer (2000, p.138) specifies that emotions are “[...] episodes of coordinated changes in several components (including at least neurophysiological activation, motor expression and subjective feeling but possibly also action tendencies and cognitive processes) in response to external or internal events of major significance to the organism”. Scherer separates emotions in terms of persistence from feelings (lower) and attitudes (higher) and confirms their role as a bridge between reactive and deliberate processes. This is consistent with a view of emotions as more mental experiences that are only weakly linked with behaviour (Schwarz and Clore 2007), leaving to more structured sensible experiences the role of shaping complex behaviour through higher cognitive processes. These are *triggered* by emotions, whose function appears to be of generating goals (Plutchik 1980; Frijda 1986),

signalling their achievement or failure (Frijda 1986; Gordon 1987) and anticipating these conditions to activate, for example, avoidance strategies (Miceli, de Rosis, and Poggi 2006). On the basis of these observations, we position emotions higher than feelings on the illocutionary gradient but lower than experiences from Munezero et al. (2014) that can be related to Scherer's *attitudes*. Emotions pertain the activation of processes belonging to the *reasoning* triangle, influencing but not directly causing action. Also, Munezero et al. (2014) highlights that emotions are preconscious, in the sense that they can become conscious through recall.

Sentiments are defined in Cattell (1940, p.6) as "[...] an acquired and relatively permanent major neuropsychic disposition to react emotionally, cognitively, and conatively toward a certain object (or situation) in a certain stable fashion, with awareness of the object and the manner of reacting". The characteristic of sentiments being lasting in time is confirmed in Murray and Morgan (1945) where they are described as an enduring disposition to respond affectively to an entity. Sentiments are more involved in decision making than emotions, thus being part of selecting behaviours. In their view of *prototypical emotional episodes*, Russell and Barrett (1999) specifies that these are strongly linked with behaviour in that they involve plans. Therefore, we do not associate them with emotions but rather with sentiments. Being more involved with decision making, we consider sentiments to be higher in the illocutionary gradient than emotions and fully involved in the execution of processes in the *reasoning* triangle which, in turn, start processes pertaining the *acting* triangle.

Opinions are powerful constructs that, while not necessarily involving an emotional component (Sokolova and Bobicev 2011) are profoundly subjective, being defined in Munezero et al. (2014, p.104) as "[...] personal interpretations of information that may or may not be emotionally charged". We position them at the top of the illocutionary gradient for this continuum to indicate that they are strongly involved with the higher forms of thought like metacognition, simulation of future scenarios, mind reading, and counterfactual reasoning (Baumeister and Masicampo 2010). All these processes pertain the *reasoning* triangle. Because of the kind of process they pertain to, opinions are also involved in the formation of moral judgements. In particular, opinions have been found to be used to fill in information gaps when producing explanations for observed/performed behaviours, producing *confabulations* (Gazzaniga and LeDoux 2013). In the specific case of people not being able to fully access the internal processes that produced a behaviour, opinions become a post-hoc reason for performing the behaviour (Haidt 2001, 2012). Specific training aimed at identifying causes for one's own behaviour and, if necessary, resist them involves accessing, processing and accepting the preconscious component igniting decision making: *emotions*. The definition of emotions as preconscious phenomena is fundamental, here, as it implies that the emotional trigger *can* be accessed through recall but, at the same time, that it is not automatically available to metacognitive processes. As a matter of fact, the instrumental relationship between sensible experiences and goals is not necessarily represented in the mind of a person (Miceli and Castelfranchi 2007). The ability to perform this kind of self-evaluation is one of the highest manifestations of human thought: philosophical thinking.

4.2 Drivenness

As explained in the Section 4.1, sensible experiences are strongly linked with goals. An important distinction has to be made concerning the type of goals different behaviours attend to. In Miceli and Castelfranchi (2014), the distinction between *weak* and *strong* purposive behaviours is linked to two different goals categories. First, *pseudo-goals*: a

"[...] functional way in which the system has been designed (by evolution, by learning, by a designer)" (Miceli and Castelfranchi 2014, p.5). Pseudo goals are also described as "[...] regulatory principles or functions concerning the system's functioning" (Miceli and Castelfranchi 2014, p.34). Second, *goals-proper*, consisting of "[...] a complex family of internal goals motivating deliberate behaviour and including wishes, desires, motives and intentions" (Miceli and Castelfranchi 2014, p.5). While these concepts are relevant, too, we will not consider them in this paper for space reasons. In the cognitive-motivational approach, *goals-proper* are internal mental representations driving the agent's behaviour after comparison with the perceived state. This results into an activity that pertains the *acting* triangle, behaviour, aimed at modifying the context to push it towards a desired, regulatory state.

Pseudo-goals are involved in regulating *goal-oriented* behaviour, which is activated automatically when relevant information is received. *Goals-proper*, instead, are involved in regulating goal-directed behaviour, based both on perceived information and on its internal representations of the anticipated results (Castelfranchi 1998). Pseudo-goals are mostly linked to the *perception* triangle in that they do not involve cognitive processes in producing behaviour.

Goals-proper are involved in the *representation* and *reasoning* triangles but their different degrees of complexity calls for a further step of discretisation for the scope of this paper. Specifically, we differentiate between *endogenous* and *exogenous* goals to account for the documented difference between intrinsic and extrinsic motivation and provide a better alignment of this continuum with the illocutionary gradient. Exogenous goals are involved in processes dealing with the *reasoning* triangle because they are meant to provide information about others' goals and, possibly, to attend them without an explicit need to alter the context for personal reasons. Endogenous goals are also involved in the *reasoning* triangle because they aim at changing the context to attend internal motives rather than to attend external changes. In an externally stable context, that is one that would not change without intervention, a *destabilising* behaviour would only happen to pursue an endogenous goal.

The type of goal that is active during behaviour performance is important and provides experiences that may be different on the sensibility continuum. As many other capabilities associated with cultural and social competence, motivation towards staying engaged in a task must come from its perceived endogenous value. If to attend exogenous goals emotions as a monitoring mechanism are sufficient to produce behaviour, that behaviour will be deprived of endogenous value because of the absence of endogenous goals, coherently with the literature predicting lower interest in activities that are triggered by external *controlling* behaviour (De Charms 2013).

Comparison Table 1 shows a comparison of the anchors in this continuum with the ones identified in the previous one. This comparison between categorical elements in the considered continua is functional to analyse the theoretical consequences of their alignment and verify that compatible anchors are linked on the same level.

4.3 Causality

Goals provide targets to reach while how to reach these goals through behaviour is a process rooted in the understanding of cause-effect relationships. These may issue different perceived *loci of causality* (Deci and Ryan 1985) in the performer, altering behaviour itself. Causal inference has attracted a significant amount of interest in the computer science community and the accomplishments of this line of research are beautifully summarised in Pearl and Mackenzie (2018). In his work, Pearl uses the *ladder*

of causation to represent two basic concepts: that a) processes pertaining to higher rungs need information coming from processes pertaining to lower rungs and b) processes on lower rungs cannot perform operations pertaining to processes on higher rungs. This is compatible with the illocutionary gradient so the anchors coming from Pearl's work fit naturally in our model. Furthermore, we adopt the general hierarchical relationship principles in all our continua. Pearl considers *seeing*, *doing* and *imagining* as keywords to the three levels *association*, *intervention* and *counterfactuals*. We adopt the same keywords as anchors, here, substituting *seeing* with *perceiving* to include a wider range of possible senses.

Perceiving is the capability of observing the environment and automatically anticipate its evolution through associative capabilities. This kind of activity pertains the *perceiving* and the *representing* triangles as it is based on fast retrieval of previous experiences with perceived stimuli and on evoking similar concepts. Abilities on this rung do not need to have an understanding of causality principles but only of recurrences, thus implying a low level of illocutionary strength. Activities involved in *perceiving* attend pseudo-goals, therefore relying on affect and feelings without the involvement of cognitive processes.

Doing is the capability of producing actions in an interventionist way to actively alter the situation. Given Austin and Searle's approach, described in the Introduction, we believe it is crucial to underline the importance of the role of *doing* in both a language philosophy model and in a computer science one. If to say something is to *do* something and *doing* something pertains to a level of intelligence going beyond association, Conversational AI using machine learning only can never truly speak because this is an activity pertaining to a superior rung in the ladder of causation. On the other hand, it is not possible to have a Conversational AI application *do* something without associative data coming from the lower level. Activities concerned with *doing* are involved in the *reasoning* and *acting* triangles and make use of causality principles to change the context, so their illocutionary strength is higher than the ones involved in *perceiving*. The ability of adopting interventionist stances towards the perceived context supports pursuing exogenous goals using emotions to monitor adjustment to the situation.

Imagining is the capability of representing future or alternative scenarios by considering multiple decision and the causal effects linked to changing these. It allows to reason on unobserved or unobservable situations, like establishing what the current context may have been if a different choice was made in the past. Activities concerned with *imagining* are mostly involved in the *reasoning* triangle but may also be used for processes producing behaviour that require self-access, future commitment and theory of mind (Premack and Woodruff 1978; Wellman 2018), thus being the strongest anchor on the illocutionary gradient on this continuum.

Goal-directed behaviours presuppose a strong link with causality understanding as it is necessary, to *do* something with them, to be able to invert the direction of causality, considering effects and going back to causes that can be created to obtain those effects as a function of goal pursuing (Pezzulo, Butz, and Castelfranchi 2008). Also, perceived locus of causality in goal pursuance is fundamental, in psychological research, concerning the effects of extrinsic motivation in the learning experience with respect to intrinsic motivation, when an activity is performed for its own sake. Any person producing an effort, either mental or physical, will perceive it as being intrinsically or extrinsically motivated and the two cannot coexist. Intrinsic and extrinsic motivation in the framework of the Self-Determination Theory (Deci and Ryan 1985; Ryan and Deci 2000) may be accounted as the result of the interaction of the *sensibility*,

drivenness and *causality* continua. Self-Determination Theory predicts the motivational value of an activity by considering its link with three basic *mental needs*:

- Autonomy: a sense of choice and volition in the regulation of behaviour;
- Competence: the sense of efficacy one has with respect to both internal and external environments;
- Relatedness: feeling connected to and cared about by others.

In general terms, intrinsic motivations provide satisfaction with respect to these needs whereas extrinsic motivations provide more *material* feedback. Rewards are a typical example of extrinsic motivation but their role “[...] can be interpreted by recipients primarily as controllers of their behaviour or, alternatively, as indicators of their competence” (Deci, Koestner, and Ryan 1999, p. 628). When rewards are perceived as means to control behaviour, however, they externalise the perceived locus of causality (De Charms 2013), thwarting satisfaction of the need of *autonomy* and reducing intrinsic motivation. When rewards are perceived as positively informational, they may provide satisfaction of the need for competence and thus enhance intrinsic motivation. Rewards, on the other hand, can take different forms and their influence on intrinsic motivation changes depending on it. *Verbal* rewards are particularly interesting to consider for the scope of this paper. In Ryan (1982), it was demonstrated that positive feedback formulated as “Excellent, you should keep up the good work” put pressure on the subjects by soliciting them in continuing to perform well and was perceived as more controlling. This was shown to reduce intrinsic motivation with respect to a second group that was simply informed they were doing better than average and suggests that the negative effect of control counterbalanced the effect of positive information. An age effect also appears to be present: the meta-analysis of previous studies presented in Deci, Koestner, and Ryan (1999) shows that although positive feedback enhanced intrinsic motivation for college students, it did not work for children. Furthermore, the same study observed that “[...] by far the most detrimental type of performance-contingent rewards - indeed, the most detrimental type of rewards — is one that is commonly used in applied settings, namely, one in which rewards are administered as a direct function of people’s performance”. Specifically, children will not express their full potential in activities that are extrinsically motivated (Ross, Karniol, and Rothstein 1976).

Comparison Table 2 shows a comparison of the anchors in this continuum with the ones identified in the previous ones.

4.4 Uncertainty

Causal effects are not always deterministic. In particular, dealing with hypothetical scenarios calls for capabilities belonging to higher rungs on the ladder of causation. This implies that dealing with uncertainty with the outcomes of producing a behaviour implies a higher illocutionary strength in producing that behaviour.

At the lowest level of the gradient, on this continuum, we position *associative learning*, which encompasses behaviours that have been either selected by evolution or learned through conditioning. They are activated as an automatic response to a stimulus and non-execution is often recognised as a malfunction. These are concerned with processes in the *perception* and *action* triangles, with a possible role in the *representing* triangle for semantic activation of relevant concepts in an associative way.

Then, we have *logical reasoning*, dealing with the ability to apply logical principles to produce *possible* courses of action to pursue a goal. Logic occupies a significant role in Artificial Intelligence and, although approaches relying entirely upon it have failed to deliver their promises, there is value in the knowledge that has been produced in past research. In our view, logical thinking serves two purposes: it manages communication repair strategies, checking for structural problems in the common ground (Clark and Schaefer 1989; Stalnaker 2002) and repairing it, and it filters out courses of action to support Bayesian reasoning in concentrating only on what is logically possible. This way, logical reasoning elaborates on associatively recovered information to filter out what is actually useful and in which possible strategies so it is concerned with the *representing* and *reasoning* triangles, activating processes belonging to the *action* triangle in unambiguous situations.

As stated by Sir Arthur Conan Doyle in his Sherlock Holmes novels, when you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth. The highest anchor on this continuum is represented by *Bayesian reasoning*. This is fundamentally different from statistical pattern matching, performed by machine learning algorithms, in that it works in a goal-directed way rather than in an goal-oriented way. However, while *associative learning* cannot do what *Bayesian reasoning* can, the latter needs the former to work, coherently with the hierarchical relationship principles from the ladder of causation. *Bayesian reasoning*, in our view, *ranks* possible courses of action considering the probability distributions of values over uncertain variables, their influence on each other and the causal impact over endogenous goals. By considering how possible actions may alter the context given the distribution of probable configurations, the most *useful* behaviour can be selected. In doing this, *Bayesian reasoning* is related with processes belonging to the *reasoning* triangle and activates processes in the *acting* triangle.

The distinction between being able to identify what is possible and what is probable has important consequences concerning sensible experiences. In the cognitive-motivational model, holding a belief about the possibility of a proposition *p* being true or false gives rise, when combined with the goal of *p* being true or false, to an Interested Anticipatory Representation (IAR): “[...] a belief-goal compound constituted by a belief about a future event *p* together with a goal that *p* or not-*p* obtains” (Miceli and Castelfranchi 2014, p.26). IARs represent a powerful motivational force in that they “[...] increase the individual’s motivation to take action and to try to realize it whenever his or her intervention is supposed to favor the attainment of the desired outcome” (Miceli and Castelfranchi 2014, p.38). Also, IARs imply an understanding of causal relationships as, being anticipatory representations, in them “[...] the causal relationship between anticipation and emotion is reversed: *the representation of a future event produces an emotional response*” (Miceli and Castelfranchi 2014, p. 26). This, characteristic, together with their direct link with interventionist capabilities, positions IARs on the second rung of the ladder of causation: *doing*. Dealing with possibilities, IARs are linked to logical reasoning, in our model.

When an IAR is combined with a expectation that *p* or not-*p* will occur, a *prediction* forms. Whether the goal is believed to be possible to achieve or expected generates profoundly different sensible experiences especially in the presence of belief invalidation through contextual evidence, leading to predictable linguistic behaviour (Domaneschi, Romero, and Braun 2017; Di Maro, Origlia, and Cutugno 2021a, 2021b). Also, expectations directly support *adaptation* (Dennett 1991), so it is important to separate them from IARs, which deal with possibilities from a cognitive-motivational point of view. Expectations are also reported to be involved in the preparation processes to decision-

making (Heckhausen and Gollwitzer 1987), so it is coherent to position them among the ones handled by *Bayesian reasoning*.

Comparison Table 3 shows a comparison of the anchors in this continuum with the ones identified in the previous ones.

4.5 Initiative

Along this continuum, we can identify a series of communicative interventions that are progressively distributed across different levels of intentionality, cognitive involvement, and social function. This spectrum ranges from minimal forms of participation to more elaborate and reflective acts. At the initial positions of the continuum, we find *backchannels* and *acknowledgments*. These are often treated as a single category of listening signals, since both express, in different ways, interactional engagement and participation in the conversational activity. For example, in Mereu, Cangemi, and Grice (2024), the term backchannel refers to brief verbal (or non-verbal) productions by a listener while the speaker holds the floor. At their core, these reactions indicate that the speaker has heard their interlocutor's utterance, but they often also show that the utterance has been understood and accepted. On the other hand, Traum (1994) uses the term acknowledgment to indicate signals that show understanding of the contribution, but not (yet) its acceptance. Similarly, Allwood, Nivre, and Ahlsén (1992) use the term feedback to refer to various types of signals across four communicative dimensions: contact, perception, understanding, and acceptance. For greater analytical clarity, and to avoid the confusion and terminological overlap arising from the complexity of the phenomenon itself, we prefer to distinguish between backchannels and acknowledgments, while recognizing that the two types of signals often coincide and that their distinction is partly conventional, thus confirming that the continuum is the suitable form to represent such phenomena.

First, we have backchannels, namely brief verbal or paraverbal signals, such as *mh* and *yes*, which indicate that the listener is following the discourse and that the communicative channel remains open. They do not introduce new content, do not take the floor, and do not necessarily signal deep understanding, but they provide minimal feedback that sustains the flow of interaction. Even in the absence of such signals, the speaker may resort to requests such as "Are you there?" or "Are you following me?", which serve to verify the continuity of the interlocutor's attentive presence.

A second level is represented by acknowledgments, or recognition acts, which indicate not only perception but also understanding of what has been said. These signals may be more explicit, such as "Ok", "I understand", or using a repetition/paraphrase of the utterance, or more implicit, such as a coherent response to a question. These acts begin to introduce a certain degree of content and show greater cognitive involvement. Again, if understanding is in doubt or not visible, the speaker may intervene with requests such as "Did you understand?" or "Is that clear?".

Proceeding along the continuum, we encounter *clarification requests* (CRs). In this case, the interlocutor does not merely signal listening or understanding but intervenes actively to resolve ambiguities, fill information gaps, or verify interpretations. These acts require a stronger cognitive and pragmatic commitment. This category includes both confirmation requests and clarification requests, used both to manage informational content (cooperation) and to negotiate the contextual frame of the interaction (negotiation). In this case, the former show a lower illocutionary strength, guided by basic communicative goals such as ensuring understanding of the received contribution, while the latter display a higher degree of illocutionary strength. Essentially, at

this stage, the presence of an (exogenous) goal is clear, but its type and its importance are not yet relevant.

Even more demanding is the act of responding to a question (QA), which constitutes one of the most structured forms of communicative cooperation. Answering a question implies taking up the other's initiative and actively contributing to the development of discourse. Illocutionary strength here is high, since the intervention has a defined purpose and directly affects the construction of shared meaning. Responses with lower illocutionary strength are those typical of adjacency pairs (Schegloff 1968; Schegloff and Sacks 1973), in which question and answer are linked by a conditional relevance relation (e.g., a request calls for acceptance or refusal).

After that, we position *proactivity*. Here, the participant does not merely react but takes the initiative: introducing new topics, anticipating needs, proposing solutions, redirecting the conversation. This is a high level of involvement, where communication becomes creative, oriented not only toward the present but also toward the future of the interaction. In Brenna, Jezek, and Magnini (2025), proactivity is defined as a collaborative-interactional phenomenon in which a participant provides their interlocutor with useful information not explicitly requested. According to the authors, proactive behaviour is: (i) self-initiated and not merely reactive (ii) innovative, since it is not strictly tied to what has already been shared in the dialogue; (iii) effectively goal-oriented, since it reflects long-term, goal-driven behaviour capable of anticipating future states and needs (Balaraman and Magnini 2020). In fact, with proactivity we approach, along other axes of the continuum, the endogenous goals pursued by the speaker employing such speech acts. In task-oriented dialogues, Brenna, Jezek, and Magnini (2025) highlight that proactivity includes a set of particular speech acts, among which inform acts (60.2%) are the most prominent, as anticipated by the definition itself. Other recurring acts include suggest, offer, request, and instruct. Finally, it is interesting to note that proactivity plays a crucial role in recovering from goal-failure situations, thereby contributing to the overall effectiveness of dialogue. Dialogues characterized by higher levels of proactivity record fewer instances of failure (Brenna, Jezek, and Magnini 2025).

At the far right end of the continuum, we find self-access capabilities. These are tools of self-reflection, useful for rationalizing decisions, organizing thoughts, analysing events, and building projections. They aim at post-hoc rationalisation of behaviour, knowledge enrichment and recent events analysis aimed at supporting commitments to the future. Although it is not necessarily expressed verbally, self-access represents a sophisticated form of linguistic activity, which fully reflects the intentionality, cognitive complexity, and organizational function of language.

Comparison Table 4 shows a comparison of the anchors in this continuum with the ones identified in the previous ones.

4.6 Performance

We define the axis of Performance as the extent to which communicative acts disclose the speaker's internal state or provide access to that of the interlocutor. Yet, the classification of speech acts has proven to be a challenging endeavour, from Austin to Searle, let alone the attempt to arrange them along a continuum of illocutionary force. In developing his framework, Searle (1976) builds on Austin's original theory of speech acts. Like Austin, he distinguishes between *illocutionary* and *perlocutionary* acts (Austin 1962), but he does not separate locutionary from illocutionary acts. Instead, he introduces the notions of *utterance acts* (the mere production of words) and *propositional*

Table 2

Hierarchical arrangement of speech acts based on a combination of engagement, commitment and initiation.

Act	Definition	Initiative	Description
Social Signals	Non-utterance as a communicative choice; can signal refusal, hesitation, or waiting	Non verbal feedback (backchannel, acknowledgment)	Minimal engagement
Assertives	Utterances that convey information or describe reality, committing the speaker to its truth	Verbal acknowledgment and inform responses	Base level of interaction: they provide factual content or feedback
Declaratives	Utterances that directly change the social or conversational state	CRs and Q&A	Acts performing social functions in interaction, like asking for clarification and responding to a question
Directives	Utterances that attempt to get the addressee to do something	Proactivity	They initiate action or information-seeking; first part of adjacency pairs with proactive intentions
Expressives	Utterances that convey the speaker's emotional state or attitude	Self-access	They add subjective/emotional response with a higher commitment
Commissives	Utterances that commit the speaker to a future action	Self-access (Promises)	They represent the highest level of speaker engagement, committing to future behaviour

acts (acts of referring and predicating). In this sense, Searle's taxonomy is not a radical departure from Austin, but rather a refinement that seeks to capture the multiple dimensions of how speech acts function. His classification is grounded on a set of analytical parameters that describe how speech acts relate to the world, to the interlocutors, and to the discourse context, namely, the illocutionary point, the direction of fit, the sincerity condition, the degree of force, the status and role of the interlocutors, the relation to the interests of the interlocutors, the position in discourse, the propositional content, and the status of verbal or non-verbal act.

Despite Searle's attempt, characterizing each act exclusively through the aforementioned traits proves to be a considerable challenge. While early frameworks treated illocutionary types as discrete categories, subsequent work has emphasized their gradient and overlapping nature. Recanati (1987) argues that illocutionary force is not a fixed property of sentence types but contextually modulated, forming a continuum between semantic content and pragmatic realization, a view further refined in Recanati (2004). Similarly, Kerbrat-Orecchioni (2001) introduces the notion of a *continuum illocutoire*, rejecting rigid taxonomies in favour of fuzzy boundaries between assertive, directive, and expressive functions. (Sbisà 2001) also develops this graded conception, proposing that illocutionary force varies in degree of strength rather than belonging to mutually exclusive categories. As a matter of fact, Sbisà (2011) explicitly notes that classification efforts inevitably end up clashing with the complexity of the illocutionary act. These perspectives converge with fuzzy-set and parametric accounts of speech acts (Kissine 2013; Allan 2013), which view each act as a composite of interacting parameters rather than a single categorical type.

Following these approaches, we adopt from Searle the notion of traits and the list of acts to be used as anchors, but, following Sbisà and Kerbrat-Orecchioni, we use them as a reference, focusing on a continuous view compatible with the illocutionary gradient. A given act, for instance, may display features typical of directives while simultaneously retaining elements of declaratives or assertives, depending on the illocutionary strength. This hierarchical arrangement is therefore adapted to the scope of this work and is outlined in Table 2.

The hierarchical structure of speech acts can be, therefore, understood in terms of increasing initiative and conversational function. At the base of the hierarchy, *Social Signals* represent the absence of verbal action. They function as minimal engagement,

often providing nonverbal feedback such as acknowledgment or backchannel signals, and serve as a communicative choice that may indicate hesitation, refusal, or waiting. Cognitively, they correspond to processes where meaning is conveyed through minimal cues without deliberate reasoning.

Moving upward, *Assertives* involve verbal acknowledgment or the provision of factual information, committing the speaker to the truth of what is expressed. These acts form the base level of active interaction, delivering content and feedback that maintain the flow of conversation. Here, the alignment is with awareness and logical reasoning: the act is grounded in perception but moves into the realm of cognition, engaging evaluative processes affirming reality.

Declaratives go further by performing social or conversational functions; they directly change the state of the interaction, for example by opening a dialogue, requesting clarification, or responding to a question. Cognitively, this requires a strategic analysis of context and authority, since the utterance creates new states of affairs. This parallels the transition into planning under uncertainty in the previous axis, where reasoning is not just about describing the world but actively transforming it. The sequence Social Signal-Assertives-Declaratives illustrates a gradual rise in engagement and effect on the interaction.

Directives initiate action or information-seeking, representing first-turn speech acts in adjacency pairs. They are goal-directed acts that imply a the capability to perform reasoning under uncertainty, given the presence of unobservable information in the interlocutors' minds. Directives assume interlocutor agency and require the speaker to anticipate responses, resonating with decision-making processes.

Expressives add an emotional or subjective layer to the interaction, communicating the speaker's feelings or attitudes with a higher level of personal involvement. These acts project inner states outward, binding cognition with affect. They map closely to the sentiment and self-access layers of the previous continuum, where emotions are not just reactive but integrated into reasoning and social interaction.

At the top of the hierarchy, *Commissives* demonstrate the greatest commitment, as the speaker undertakes responsibility for future actions, closing the interaction loop and representing the highest level of engagement. This involves endogenous goals, self-reflection, and complex intention formation. Commissives resonate with the processes of Bayesian reasoning and prospective reflection in the other axis: the speaker evaluates uncertainty, projects possible outcomes, and commits to one of them, thereby engaging imagination and long-term planning. This structure reflects a continuum of conversational involvement, from passive or factual contributions to acts that actively shape the dialogue and commit the speaker to future actions.

Anchors on this continuum are better described as aspects that may be added to speech acts depending on the level of illocutionary strength needed to implement them. This continuum is better understood as a dynamic combination of parameters that, when assembled in different ways, define the illocutionary strength of a given act and its role in interactional performance. This way, the hierarchy ranges from minimal participation (Social Signals) to maximal engagement and responsibility (Commissives), while also capturing the varying roles that acts play in maintaining, initiating, or transforming the conversational exchange. The Performance continuum deals with surface traces of deeper dynamics represented in the other axes. This has important implications in simulating a *raison d'exprimer* in machines: the ability to use those traits depends on the possibility to simulate the corresponding levels of illocutionary force so that, for example, a machine equipped with logical representation capabilities can use Assertives

and Declaratives but not Directives and Expressives, for which Bayesian reasoning would be needed.

Comparison Table 5 shows a comparison of the anchors in this continuum with the ones identified in the previous ones.

5. Raison d'exprimer

We now switch to describing how the aligned continua relate to our view of a Conversational AI architecture. First of all, the use of illocutionary gradient as a common dimension provides us with a way to rank processes involved in Conversational AI and the kind of linguistic output they produce to identify *why* a technological system said something. This resonates with cognitive models of decision-making emphasizing how agents operate in different modes depending on the situation. From completely automatic behaviour to decision making using heuristic reasoning, affective cues, and probabilistic expectations over uncertain variables (Kahneman and Tversky 1979). Higher cognitive capabilities enabling mental simulation, counterfactual thinking, and probabilistic inference, align with Bayesian approaches to cognition that conceptualize the mind as a predictive engine operating through hypothesis testing and belief updating (Griffiths and Tenenbaum 2006).

According to Lorini, Marzo, and Castelfranchi (2005), Miceli and Castelfranchi (2014), a motivating goal of a certain behaviour is one that is perceived as a necessary and sufficient reason for performing that behaviour. We argue that, in the context of Conversational AI, this corresponds to *raison d'exprimer*. The aligned continua and the corresponding capabilities of a Conversational AI architecture are shown together in Figure 9, discussed in the rest of this Section.

First, we account for completely automatic behaviour. We use the concept of *neuroception* (Porges 2004) to refer to hard wired processes designed to distinguish whether situations are safe, dangerous, or life threatening, producing *action* without supervision from higher cognitive processes. This mechanism is the root of our model as it may even disable social behaviour when unsafe conditions are detected. Neuroception is about *presence* and *perceiving*, activating *associative learning* regulated by *affect* to handle unconscious *pseudo-goals*. For the case of communication, these are related to social signals like blushing, posture mimicry and micro-facial expressions.

Perception consists of hard-wired capabilities to pre-process input to make it interpretable. While perceiving communicative signals, continuous prediction of the next units in observed behaviour is performed, a process called *sensoryprediction* that has been shown to help prepare reactions to stimuli before they are detected (SanMiguel et al. 2013). *Perception* informs *sensoryprediction* which, through associative learning, anticipates incoming stimuli to activate *pre-actions*. These represent, in our model, the impulse to act, subsequently converted in actual communicative signals, constituting the *action*. Generative AI models are positioned here: associative learning implements *sensoryprediction* and the internal configuration of the network weights constitutes the *pre-action*. Words outputted by the model are, instead, the *action*. These models, going beyond purely wired behaviour to add learning abilities, have an adaptive internal configuration that can be assimilated to a simulation of *feelings* but still produce *associative learning* actions to attend the *pseudo-goal* of completing the text. In producing actions this way, their output has been found to match at least one formal definition of *bullshit* (Hicks, Humphries, and Slater 2024). This position also matches Pearl and Mackenzie (2018) in stating they do not go beyond *perceiving*. This also matches the neurobiology of reward-related learning, identifying most of the brain activity related to it in the

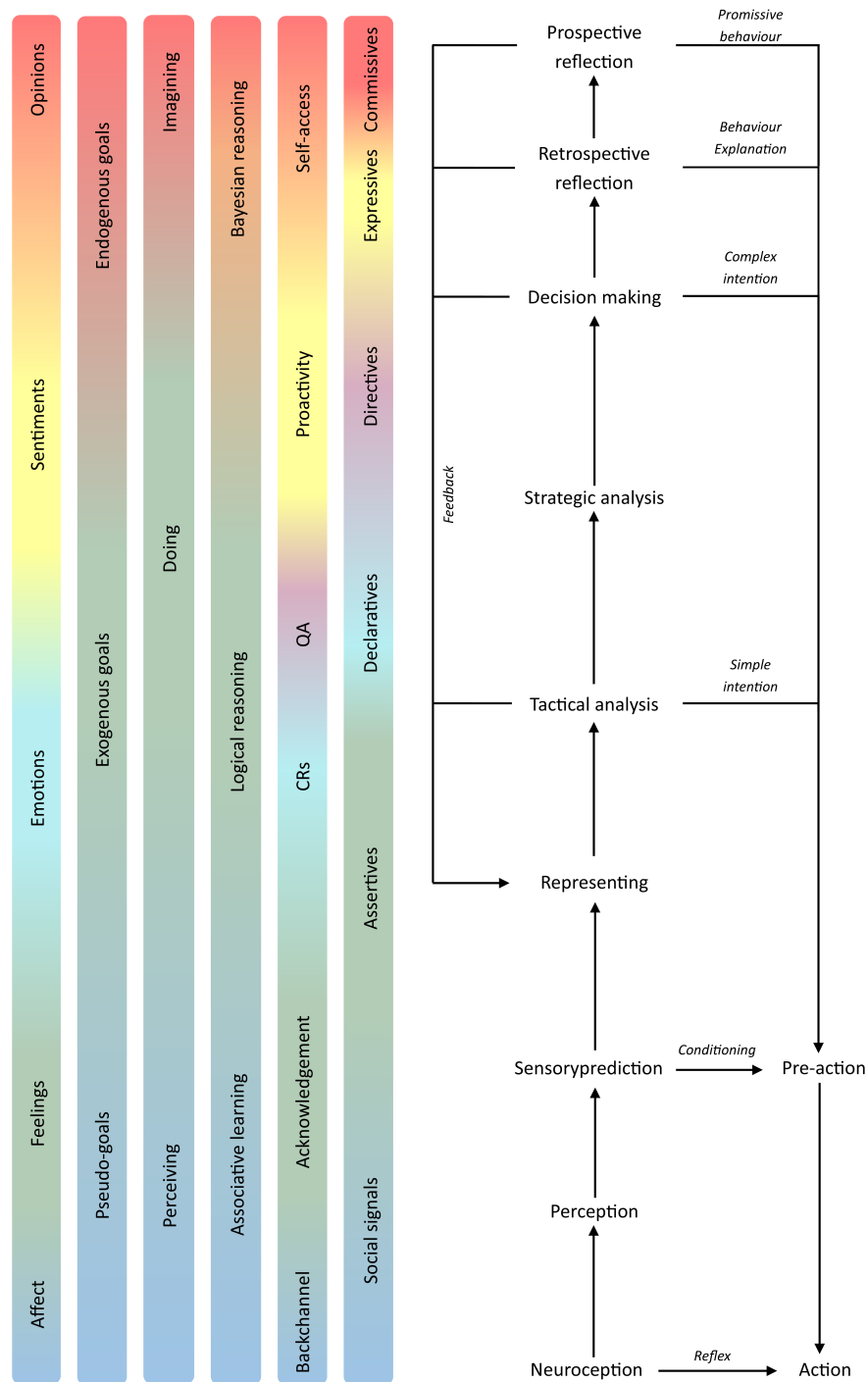


Figure 9

A representation of a cognitively inspired architecture for Conversational AI building from neuroception to higher capabilities in parallel with the considered continua related to communication. On each continuum, *anchors* are reported.

mesencephalic area (Galaj and Ranaldi 2021), outside the control of the neocortex. While it is possible, in the brain, for the neocortex to inhibit elicited conditioned reflexes, in modern LLM architectures this role is delegated to rule-based or prompt-based *guardrails*. These, being either external to the system or part of the stimulus, do not represent decision-making. As a consequence, LLMs are simply *present* with a minimal illocutionary force. It would be a mistake to infer from the statistical character of these systems that they lack internal structure. Work in mechanistic interpretability has established the contrary: training gives rise to identifiable internal mechanisms, such as the induction heads that support in-context copying and pattern continuation (Olsson et al. 2022), and to approximately monosemantic features that can be isolated within the activation space of large models (Templeton et al. 2024). On Dennett's criterion of *real patterns* (Dennett 1991), according to which a pattern is real if positing it affords genuine predictive purchase, these structures are unquestionably real. We do not dispute their reality, and our argument does not require us to. The internal structures revealed by interpretability are, in the same way, real patterns organised around a particular objective, the minimisation of predictive error over a corpus, which we have termed a pseudo-goal. Our claim is therefore not that these patterns are illusory or merely simulated, an opposition we explicitly abandon, but that they are organised around an exogenous objective rather than around endogenous goals generated by the system's own value system. A bare language model does not even adopt the goal a prompt expresses: it conditions its output on the prompt yet keeps pursuing its pseudo-goal, generating the most plausible continuation, independently of the goal the prompt states. The distinction we require is therefore not between genuine and fake intentionality but among organising principles: those that are simply wired, the pseudo-goal of plausibility, those that adapt to externally supplied goals, and those that adapt to endogenously generated ones, the last being where the organising principle becomes a *raison d'exprimer*.

If a system's behaviour is reliably predictable by attributing goals to it, then on Dennett's own terms it has goals, and to ask whether those goals are endogenous or not would be to demand something over and above the predictive pattern. To avoid this, the distinction between exogenous and endogenous organisation, in our view, does not sit behind the patterns: it is itself a higher-order pattern, manifest in behaviour and supporting its own predictions. A system whose goals are exogenously supplied and a system whose goals are endogenously generated differ in ways that are both observable and compressible: the latter pursues goals across contexts without being prompted to, resists substitution of one goal for another, sacrifices local reward to preserve a standing commitment, and exhibits the asymmetry between merely entertaining a goal and undertaking it. These regularities are as real, by Dennett's own criterion, as the lower-order patterns they organise. The exogenous/endogenous distinction is thus not a metaphysical residue but a real pattern in how a system's goal-patterns are generated and maintained, a pattern that current systems, organised wholly around the corpus-plausibility objective, do not instantiate. Our position stays within the intentional stance of (Dennett 1987): both kinds of system have real intentional patterns but what differs, which is what the illocutionary gradient measures, is the organisation of those patterns.

A practical example of this organisation is visible in how such systems are deployed. The representational content they appear to possess is routinely supplied from outside, retrieval-augmented generation grounds their output in an external store, without which their tendency to fabricate rises, and the monitoring of that output is also externalised, in downstream classifiers that intercept unsafe or false content the system itself emits without check. A system organised around a goal of its own, and protective of it, would detect and correct its own deviations: the need for an external

monitor to register that it has gone astray indicates that the goal-protective loop is not closed within it. Belief, goal, and appraisal are in this sense each supplied from outside, by the store, the prompt, and the classifier respectively, because internally the system is organised around plausibility rather than around a goal proper it represents and defends. In linguistic terms, the pragmatic value of the content such systems generate is inferior to its syntactic and semantic value, as one expects of a system that masters the lower levels of language while lacking the goal-directed level on which pragmatics depends.

Finally, the gradient disciplines how these interpretability results are to be read. Neural networks lie low on the illocutionary continuum is argued here on independent grounds: given that placement, the explanation of patterns such as induction heads and monosemantic features should be sought among the correspondingly low-level constructs of adjacent fields. Ortony's notion of *proto-affect*, a reactive assignment of value to stimuli, below the level of belief and goal (Ortony, Norman, and Revelle 2005), sits at exactly the level the gradient predicts and is compatible with the observed patterns; and *proto-affect*, adopted by (Miceli and Castelfranchi 2014) in distinguishing *proto-fear* from fear proper and aligned with the distinction (Munezero et al. 2014) draw between affect and feeling on the one hand and emotion on the other, is itself eligible as a real pattern and offers a valid reading of the function these patterns serve. Reading the same structures as beliefs and intentions is a permissible Dennettian inference. We accept it, and advance the *proto-affect* reading as an alternative that is equally compatible with the intentional stance with the advantage of being motivated by the illocutionary gradient. We do not claim it is the uniquely correct interpretation: only that it has theoretical grounding, through the gradient, in identifiable portions of the literature that align with the other characteristics of the systems at hand. The example shows both how readily neural firing patterns are read as beliefs or intentions and how the gradient helps locate a reading at the level it assigns these systems, which is also the framework's constructive role, coordinating interpretation across fields by level.

To improve the performance of LLMs, Retrieval Augmented Generation techniques have been proposed. In these systems, prompts are automatically enriched with relevant information coming from external sources, typically databases. The separation between language generators and knowledge repositories equipped with multiple retrieval and processing mechanisms motivates the *representing* ability. This enables *awareness* by providing an explicit representation of the present context and the ability to manipulate these representations beyond simple association. In particular, the presence of a formal representation outside associative mechanisms enables two relevant types of reasoning: forward and backward chaining. Forward chaining is a data-driven reasoning approach approximating a breadth-first exploration of a solutions tree while backward chaining is a goal-driven one approximating a depth-first exploration of a solutions tree. In general, forward chaining is good at producing consequential information from initially available data, while backward chaining is good at producing plans to achieve desired objectives. Modern representations based on knowledge graphs are particularly precious as a connecting point (Di Maro et al. 2025) between *associative learning* attending *pseudo-goals* and the first form of goal-directed reasoning, *tactical analysis*, using *logical reasoning* to start *doing* things in a performative way. This kind of representation supports both chaining techniques. Forward chaining, being data-driven, can be implemented using graph data science techniques, enriching collected knowledge by making latent information emerge, attending the pseudo-goal of collecting information about the context. Backward chaining can be implemented using classic logical inference aimed at demonstrating the attainability of a goal-proper by generating solution trees. Backward

chaining is activated by *emotions* whenever the observed situation is combined with previous beliefs and current goals. Expectancy violation, in particular, is documented to support causal learning (De Houwer, Beckers, and Glautier 2002), which we may interpret as a stimulus to memorise current contextual information, an emotional episode, to help better derive causes of prediction failures during later retrospective reasoning at higher levels.

The type of goal attainable at this level is limited to *exogenous* ones, so that only goals having an external locus of causality can be pursued. In communication, this includes the ability to repair the channel using CRs and provide *question answering* about factual knowledge. Concerning the former, when asking a CR, the system is attending the exogenous goal, given by the interlocutor, to receive a message and do what is told. The driver, in this case, is entirely extrinsic. The system, at this level, can use language to *do* things in a goal-directed way but has no capability of generating these goals in absence of external input. These systems are good at *tactical analysis* in the sense that they support the achievement of material, short-term exogenous goals but have no self-determining, long-term endogenous ones. Agentic and neurosymbolic architectures combine rule-based and associative statistical models, positioning themselves on this level although the variety of the approaches calls for caution⁷. Approaches centred, for example, on tool-calling abilities from LLMs fall back to the first rung of the ladder of causation. This is because, to be considered an indication of second rung abilities, tool use, including language, must be “[...] intentional and not just accidental or copied from ancestors” (Pearl and Mackenzie 2018, p.27), in other words, a mere result of training/conditioning. This is a functional criterion, not a biological one. It marks the distinction between associative behaviour (the first rung of Pearl’s ladder) and behaviour grounded in a causal model of why a means brings about its effect (the higher rungs). Behaviours that appear superficially identical may differ in the reason why they were produced, which is part of the functionalist approach.

Goal-directedness is itself a functional property: a behaviour is goal-directed when the goal is the variable the behaviour is organised to achieve, so that means vary to reach it and deviations are corrected. In the closing of a gap by plausible continuation, the variable the behaviour is organised to bring about is plausibility. When a system produces language that closes an information gap, the closing of the gap is a downstream coincidence rather than what the behaviour tracks, and where plausibility and gap-closure diverge, the system follows plausibility and leaves the gap open. Indeed the system’s goal-oriented surface behaviour is to reach the end-of-sequence token along the most plausible path, so that by its own stopping criterion the pseudo-goal is attained whenever that criterion is matched. The indifference to truth that (Hicks, Humphries, and Slater 2024) characterise as *bullshit* follows from this mechanism, and is what is commonly called *hallucination*; this, is, however, an attributed phenomenon, too. The system is, so to speak, content with its token stream, and having had no goal other than producing the most probable continuation, it does not, strictly, hallucinate at all. The behaviour is therefore not, on the functional analysis itself, goal-directed toward closing the gap. Without considering the internal goal, it is only *behaviourally* indistinguishable from a goal-directed act.

⁷ Commentary on production coding agents has noted that critical decision logic may be realised in classical symbolic code rather than delegated to the language model (Marcus 2026); we cite this as illustration only, the underlying point being the architectural one that decision-making is externalised from the model.

Logical reasoning is not only functional to attend simple exogenous goals. It is also the basis upon which capabilities directed at attending *endogenous goals* are founded. In Section 4.4, we highlighted the difference between beliefs of possibility and beliefs of probability, constituting the difference between IARs and predictions. Logical thinking, by applying backward chaining directed towards endogenous goals, lists possible ways to achieve that goal. Those strategies are subject to evaluation against personal gain, an important difference w.r.t. material gain. Considering the study on intrinsic motivation presented in Ross, Karniol, and Rothstein (1976), children who were promised a reward “if they would draw some pictures” were compared to children who were not. The first group produced more pictures than the second one but their quality was inferior. Moreover, after being left alone, rewarded children did not continue drawing while the others did, indicating the presence of endogenous goals. The long-term persistence of these goals matches the definition of *sentiments* as a motivating element w.r.t. short lived *emotions* and makes *proactive* stances possible through the support of *strategic analysis*.

Tactical and strategic analysis use backward chaining to produce possible contexts and action plans implementing reasoning under assumptions. Ranking solutions to *endogenous goals* has a different nature than the one used for *exogenous goals*. In the former, consequences concerning self-determination should be taken into account while, in the latter, the more basic principle of minimal effort (Hull 1943) may be considered. Evaluating possible consequences over long-term goals requires dealing with a more complex network of causal relationships and with higher instability concerning available information. *Decision making*, in our architecture, uses *Bayesian reasoning* to estimate the most probable configuration of uncertain variables describing the context and the corresponding utility of possible courses of action towards the achievement of *endogenous goals*. Computational models that efficiently implement this are Influence Diagrams (Howard and Matheson 2005), an extension of Bayesian Networks explicitly considering interventions and outcomes while performing reasoning under uncertainty (Biedermann and Taroni 2022).

The capability of attending *endogenous goals* is necessary but not sufficient to shift from *awareness* to *consciousness* as this necessitates *self-access* capabilities for *retrospective* and *prospective reflection*. The former is the process of generating and evaluating alternative outcomes to the one observed, given previously unavailable information, while the latter is the process of generating new principles to use when generating and evaluating possible courses of action. Both are related to *imagining* and enable *behaviour explanation* and *promissive behaviour* by allowing the system to access and analyse its internal functioning. *Self-access* activities generate and are influenced by *opinions*, constituting metacognitive abilities. Bayesian reasoning is involved in these activities, consistently with indications that it predicts the outcomes of human causal inference in backwards blocking (Benton and Rakison 2023) better than associative models. Backwards blocking consists of causal relationships discarding as a result of retrospective analysis (Wasserman and Berglan 1998). Emotional memories fixed at a lower level are evaluated at this level, consistently with the cognitive-motivational model being “[...] a feedback theory of emotions that assigns them the role of fostering a retrospective analysis of recent events that is functional to prospective behaviour” (Miceli and Castelfranchi 2014, p.18). This delayed analysis of memorised emotional episodes is consistent with *emotions* being pre-conscious. It also makes the system prone to confabulations, which emerge during post-hoc sense-making processes when emotions recall fails. *Opinions* influence both behaviour explanation and promissive behaviour. In the first case, confabulations may be produced in an attempt to fill gaps in explaining behaviour so that the expla-

nation is consistent with an existing opinions system. In the second one, promissive behaviour implies a moral obligation to comply.

The way these capabilities are orchestrated defines a hierarchy of increasingly complex abilities implying that the system can represent more and more complex relationships between itself and the situational context. The capability of modifying the context in a performative way, together with the type of pursued goal, defines the illocutionary strength of the produced speech acts. As suggested in (Di Maro et al. 2025), a hierarchical organisation of conversational abilities may be represented using behaviour trees, a popular AI model in the gaming industry and in robotics (Iovino et al. 2022) that implicitly codes tasks priority. Specifically, Di Maro et al. (2025) suggests that high-priority tasks in such a model deal with low-level communicative necessities and low-priority ones deal with communicative strategies aimed at achieving contextual desirability. Lastly, (Di Maro et al. 2025) links this hierarchy to illocutionary strength, too, so an interesting consequence is that the priority of communicative acts is inversely proportional to their illocutionary strength.

6. Conclusions

The modern landscape of Conversational AI is intrinsically cross-disciplinary and is not limited to computer science only, but encompasses fields like linguistics and cognitive science. In this situation, it may be hard for scholars to orient themselves in a context where terminological issues and multiple points of view co-exist. In this paper, we have presented a proposal to organise the available knowledge both from *horizontal* and *vertical* points-of-view. The horizontal dimension is formalised by the hypertriangle of communication, describing how different moments in communication dynamics blend into each other. The hypertriangle also serves as a compass to explore the vertical dimension provided by the illocutionary gradient. This is designed to represent a common alignment dimension between multiple continua to foster comparisons and descriptive analyses. The continua considered in this paper are not meant to be exhaustive. However, having shown how it is possible to align six cross-disciplinary continua around this common dimension, we propose the use of the illocutionary gradient as a reference dimension to provide further alignments with other continua in future research. The advantage provided by this approach consists in the automatic comparison of new continua with the ones considered here to support critical analyses of the relationships among different disciplines concerning with Conversational AI. While we remain far from attributing to machines a true *raison d'être*, that is, a purpose or intrinsic reason for existence understood as self-ascribed, rather than externally assigned by human designers who define a system's functional goals, we believe this contribution provides a reference framework around which to build a formal view of *raison d'exprimer*.

The framework set out here is, at bottom, a philosophical one. The illocutionary gradient and the hypertriangle of communication are not merely descriptive instruments; they are a way of grounding computational accounts of dialogue in principled distinctions concerning intention, awareness, and meaning, distinctions that a purely mechanistic or market-driven account of these systems is apt to flatten. We therefore close not by resolving the hardest questions these distinctions raise, the relation between awareness and consciousness foremost among them, but by marking them as the natural site of interdisciplinary work. We regard such engagement as necessary rather than optional: if the field is to progress without mistaking what can be marketed for what has been understood, the theoretical questions concerning the nature of thought and of intentional behaviour cannot be set aside. Our aim in raising these questions at

the close is therefore not to settle them but to offer the gradient and the hypertriangle to philosophers, to AI researchers and other scientists involved in from multiple fields as a shared dimension in which they can be posed precisely.

The framework is theoretical, but it is not idle. Its architecture was anticipated and partially tested in our earlier work (Di Maro et al. 2025); a working implementation now exists, and its full deployment within an embodied conversational agent is in progress, with the complete computational account given in companion work in preparation. We indicate here how its components are intended to interact, and how the architecture bears on two classical obstacles to integrating symbolic and sub-symbolic processing. The processing levels distinguished by the framework map onto a pipeline in which each level is realised by a technology suited to it: perception by associative models that map continuous input to structured representations; representation by a graph that maintains the dialogue state; logical reasoning by a symbolic inference layer; reasoning under uncertainty by a probabilistic decision layer; and action by generation conditioned on the selected communicative act. The levels are connected by well-defined interfaces, so that the output of each is the input of the next.

Acknowledgments

This work is supported by the Supporting Patients with Embodied Conversational Interfaces and Argumentative Language (SPECIAL) project, funded by the University of Naples on the “Fondi per la Ricerca di Ateneo” (FRA) program (CUP: E65F22000050001).

The authors would like to thank Prof. Emanuela Campisi for the relevant comments and discussions during the development of this work.

References

- Ajzen, Icek and Martin Fishbein. 1980. *Understanding attitudes and predicting social behavior*. Prentice-Hall, Englewood Cliffs, NJ.
- Allan, Keith. 2013. *Linguistic Meaning*, volume 1 of *Routledge Library Editions: Linguistics A: General Linguistics*. Routledge.
- Allwood, Jens, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26.
- Austin, John L. 1962. *How to Do Things with Words*. The William James Lectures. Oxford University Press, Oxford.
- Balaraman, Vevake and Bernardo Magnini. 2020. Pro-active systems and influenceable users: Simulating pro-activity in task-oriented dialogues. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2020 – WatchDial)*, Waltham, MA, USA (virtual), July. SEMDIAL.
- Batson, C. Daniel, Laura L. Shaw, and Kathryn C. Oleson. 1992. Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions. In Margaret S. Clark, editor, *Emotion*, volume 13 of *Review of Personality and Social Psychology*. SAGE Publications, Newbury Park, CA, pages 294–326.
- Baumeister, Roy F. and E. J. Masicampo. 2010. Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review*, 117(3):945–971.
- Benton, Deon T. and David H. Rakison. 2023. Associative learning or bayesian inference? Revisiting backwards blocking reasoning in adults. *Cognition*, 241:105626.
- Biedermann, Alex and Franco Taroni. 2022. Bayesian networks and influence diagrams. In Max M. Houck, editor, *Encyclopedia of Forensic Sciences*. Elsevier, Amsterdam, 3 edition.
- Braitenberg, Valentino. 1986. *Vehicles: Experiments in synthetic psychology*. MIT press.
- Brenna, Sofia, Elisabetta Jezek, and Bernardo Magnini. 2025. Investigating proactivity in task-oriented dialogues. *Dialogue & Discourse*, 16(1):31–67.

- Burgoon, Judee K., Nadia Magnenat-Thalmann, Maja Pantic, and Alessandro Vinciarelli, editors. 2017. *Social Signal Processing*. Cambridge University Press, Cambridge.
- Castelfranchi, Cristiano. 1998. Modelling social action for ai agents. *Artificial intelligence*, 103(1-2):157–182.
- Catford, John Cunnison. 1965. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford University Press, London.
- Cattell, Raymond B. 1940. Sentiment or attitude? The core of a terminology problem in personality research. *Journal of Personality*, 9(1):6–17.
- Clark, Herbert H. and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13(2):259–294.
- De Charms, Richard. 2013. *Personal causation: The internal affective determinants of behavior*. Routledge.
- De Houwer, Jan, Tom Beckers, and Steven Glautier. 2002. Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology: Section A*, 55(3):965–985.
- de Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Payot, Paris.
- Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6):627–668.
- Deci, Edward L. and Richard M. Ryan. 1985. The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, 19(2):109–134.
- Dennett, Daniel C. 1987. *The Intentional Stance*. MIT Press.
- Dennett, Daniel C. 1991. *Consciousness explained*. Little Brown and Co.
- Di Maro, Maria, Martina Di Bratto, Sabrina Mennella, Antonio Origlia, and Francesco Cutugno. 2025. Argumentation in recommender dialogue agents (ARDA): An unexpected journey from pragmatics to conversational agents. *Open Linguistics*, 11(1).
- Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021a. Cutting melted butter? Common ground inconsistencies management in dialogue systems using graph databases. *IJCoL. Italian Journal of Computational Linguistics*, 7(1-2):157–190.
- Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021b. PolarExpress: Polar question forms expressing bias-evidence conflicts in Italian. *International Journal of Linguistics*, 13(4):14–35.
- Domaneschi, Filippo, Maribel Romero, and Bettina Braun. 2017. Bias in polar questions: Evidence from English and German production experiments. *Glossa: A Journal of General Linguistics*, 2(26):1–28.
- Ekman, Paul. 1999. Basic emotions. In Tim Dalgleish and Mick J. Power, editors, *Handbook of Cognition and Emotion*. John Wiley & Sons, Chichester, pages 45–60.
- Federmeier, Kara D. and Marta Kutas. 1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, 41(4):469–495.
- Fontaine, Johnny R. J., Klaus R. Scherer, Etienne B. Roesch, and Phoebe C. Ellsworth. 2007. The world of emotions is not two-dimensional. *Psychological science*, 18(12):1050–1057.
- Frijda, Nico H. 1986. *The emotions*. Cambridge University Press.
- Galaj, Ewa and Robert Ranaldi. 2021. Neurobiology of reward-related learning. *Neuroscience & Biobehavioral Reviews*, 124:224–234.
- Gazzaniga, Michael S. and Joseph E. LeDoux. 2013. *The Integrated Mind*. Springer, New York.
- Gibson, James J. 2014. *The ecological approach to visual perception: classic edition*. Psychology press.
- Gordon, Robert Morris. 1987. *The structure of emotions: Investigations in cognitive philosophy*. Cambridge University Press.
- Griffiths, Thomas L. and Joshua B. Tenenbaum. 2006. Optimal predictions in everyday cognition. *Psychological science*, 17(9):767–773.
- Haidt, Jonathan. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4):814–834.
- Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Heckhausen, Heinz and Peter M. Gollwitzer. 1987. Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and emotion*, 11(2):101–120.
- Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. ChatGPT is bullshit. *Ethics and Information Technology*, 26(2):38, June.
- Hjelmslev, Louis and Francis J. Whitfield. 2024. Lectures on a theory of language. In Lorenzo Cigana, editor, *Lectures on Language Theory 1942–1943*. John Benjamins, Amsterdam, pages

- 21–213.
- Howard, Ronald A. and James E. Matheson. 2005. Influence diagrams. *Decision Analysis*, 2(3):127–143.
- Hull, Clark L. 1943. *Principles of Behavior: An Introduction to Behavior Theory*. Appleton-Century, New York.
- Huron, David. 2008. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, Cambridge, MA.
- Iovino, Matteo, Edvards Scukins, Jonathan Styrud, Petter Ögren, and Christian Smith. 2022. A survey of behavior trees in robotics and AI. *Robotics and Autonomous Systems*, 154:104096.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.
- Kahneman, Daniel and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.
- Kerbrat-Orecchioni, Catherine. 2001. *Les actes de langage dans le discours : théorie et fonctionnement*. Fac. Linguistique. Nathan Université, Paris.
- Kissine, Mikhail. 2013. *From utterances to speech acts*. Cambridge University Press.
- Kleinginna, Jr., Paul R. and Anne M. Kleinginna. 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4):345–379.
- Lorini, Emiliano, Francesca Marzo, and Cristiano Castelfranchi. 2005. A cognitive model of altruistic mind. In *Advances in Cognitive Economics*. NBU Press, Sofia, pages 282–293.
- Loru, Edoardo, Jacopo Nudo, Niccolò Di Marco, Alessandro Santirocchi, Roberto Atzeni, Matteo Cinelli, Vincenzo Cestari, Clelia Rossi-Arnaud, and Walter Quattrocchi. 2025. The simulation of judgment in llms. *Proceedings of the National Academy of Sciences*, 122(42):e2518443122.
- Marcus, Gary. 2026. The biggest advance in ai since the llm. Marcus on AI (Substack).
- Massumi, Brian. 2021. *Parables for the Virtual: Movement, Affect, Sensation*. Duke University Press, Durham, NC, 20th anniversary edition.
- Mereu, Daniela, Francesco Cangemi, and Martine Grice. 2024. Backchannels are not always very short utterances. The case of italian multi-unit backchannels. *Journal of Pragmatics*, 228:1–16.
- Miceli, Maria and Cristiano Castelfranchi. 2007. The envious mind. *Cognition and emotion*, 21(3):449–479.
- Miceli, Maria and Cristiano Castelfranchi. 2014. *Expectancy and emotion*. OUP Oxford.
- Miceli, Maria, Fiorella de Rosis, and Isabella Poggi. 2006. Emotional and non-emotional persuasion. *Applied Artificial Intelligence*, 20(10):849–879.
- Munezero, Myriam, Calkin Suero Montero, Erkki Sutinen, and John Pajunen. 2014. Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE transactions on affective computing*, 5(2):101–111.
- Murray, Henry Alexander and Christiana D. Morgan. 1945. A clinical study of sentiments (I & II). *Genetic Psychology Monographs*, 32:3–149.
- Ogden, Charles Kay and Ivor Armstrong Richards. 1927. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Kegan Paul, Trench, Trubner & Co., London, 2 edition.
- Olsson, Catherine, Nelson Elhage, Neel Nanda, Nicholas Joseph, Nova DasSarma, Tom Henighan, Ben Mann, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Scott Johnston, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, and Chris Olah. 2022. In-context learning and induction heads. Transformer Circuits Thread.
- Ortony, Andrew, Donald A. Norman, and William Revelle. 2005. Affect and proto-affect in effective functioning. In Jean-Marc Fellous and Michael A. Arbib, editors, *Who Needs Emotions? The Brain Meets the Robot*. Oxford University Press, pages 173–202.
- Pearl, Judea and Dana Mackenzie. 2018. *The book of Why*. Basic Books.
- Pezzulo, Giovanni, Martin V. Butz, and Cristiano Castelfranchi. 2008. The anticipatory approach: definitions and taxonomies. In *The challenge of anticipation: A unifying framework for the analysis and Design of Artificial Cognitive Systems*. Springer, pages 23–43.
- Plutchik, Robert. 1980. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Emotion: Theory, Research, and Experience*, volume 1. Academic Press, New York, pages 3–33.
- Poeppel, David and M. Florencia Assaneo. 2020. Speech rhythms and their neural foundations. *Nature reviews neuroscience*, 21(6):322–334.

- Porges, Stephen W. 2004. Neuroception: A subconscious system for detecting threats and safety. *Zero to Three*, 24(5):19–24.
- Premack, David and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526.
- Recanati, François. 1987. *Meaning and Force: The Pragmatics of Performative Utterances*. Cambridge University Press, Cambridge.
- Recanati, François. 2004. *Literal meaning*. Cambridge University Press.
- Reisenzein, Rainer. 2009a. Emotional experience in the computational belief–desire theory of emotion. *Emotion Review*, 1(3):214–222.
- Reisenzein, Rainer. 2009b. Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, 10(1):6–20.
- Ross, Michael, Rachel Karniol, and Mitch Rothstein. 1976. Reward contingency and intrinsic motivation in children: A test of the delay of gratification hypothesis. *Journal of Personality and Social Psychology*, 33(4):442–449.
- Russell, James A. and Lisa Feldman Barrett. 1999. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5):805–819.
- Ryan, Richard M. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3):450–461.
- Ryan, Richard M. and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78.
- SanMiguel, Iria, Andreas Widmann, Alexandra Bendixen, Nelson Trujillo-Barreto, and Erich Schröger. 2013. Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience*, 33(20):8633–8639.
- Sbisà, Marina. 2001. Illocutionary force and degrees of strength in language use. *Journal of pragmatics*, 33(12):1791–1814.
- Sbisà, Marina. 2011. Teoria degli atti linguistici: appunti per una storia. Manoscritto, Università degli Studi di Trieste.
- Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist*, 70(6):1075–1095.
- Schegloff, Emanuel A. and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.
- Scherer, Klaus R. 2000. Psychological models of emotion. In Joan C. Borod, editor, *The Neuropsychology of Emotion*. Oxford University Press, Oxford, pages 137–162.
- Scherer, Klaus R. 2001. Appraisal considered as a process of multilevel sequential checking. In Klaus R. Scherer, Angela Schorr, and Tom Johnstone, editors, *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press, Oxford, pages 92–120.
- Schwarz, Norbert and Gerald L. Clore. 2007. Feelings and phenomenal experiences. In Arie W. Kruglanski and E. Tory Higgins, editors, *Social Psychology: Handbook of Basic Principles*. Guilford Press, New York, 2 edition, pages 385–407.
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society*, 5(1):1–23.
- Searle, John R. 2003. *Rationality in Action*. MIT Press, Cambridge, MA.
- Shouse, Eric. 2005. Feeling, emotion, affect. *Media/Culture Journal*, 8(6).
- Snyder, Hannah. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339.
- Sokolova, Marina and Victoria Bobicev. 2011. Sentiments and opinions in health-related web messages. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2011)*, pages 132–139, Hissar, Bulgaria, September. Association for Computational Linguistics.
- Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy*, 25(5–6):701–721.
- Stewart, Neil, Nick Chater, and Gordon D. A. Brown. 2006. Decision by sampling. *Cognitive psychology*, 53(1):1–26.
- Templeton, Adly, Tom Conerly, Jonathan Marcus, Jack Lindsey, Trenton Bricken, Brian Chen, Adam Pearce, Craig Citro, Emmanuel Ameisen, Andy Jones, Hoagy Cunningham, Nicholas L. Turner, Callum McDougall, Monte MacDiarmid, Alex Tamkin, Esin Durmus, Tristan Hume, Francesco Mosconi, C. Daniel Freeman, Theodore R. Sumers, Edward Rees, Joshua Batson, Adam Jermyrn, Shan Carter, Chris Olah, and Tom Henighan. 2024. Scaling monosemanticity: Extracting interpretable features from Claude 3 Sonnet. Transformer Circuits Thread.
- Torraco, Richard J. 2005. Writing integrative literature reviews: Guidelines and examples. *Human resource development review*, 4(3):356–367.

- Traum, David R. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester, Rochester, NY. Also published as Technical Report 545, Department of Computer Science.
- Tulving, Endel. 1972. Episodic and semantic memory. In Endel Tulving and W. Donaldson, editors, *Organization of Memory*. Academic Press, New York, pages 381–403.
- Vinciarelli, Alessandro, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759.
- Wasserman, Edward A. and Lyndon R. Berglan. 1998. Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *The Quarterly Journal of Experimental Psychology: Section B*, 51(2):121–138.
- Wellman, Henry M. 2018. Theory of mind: The state of the art. *European Journal of Developmental Psychology*, 15(6):728–755.
- Wertheimer, Max. 1938. Laws of organization in perceptual forms. In Willis D. Ellis, editor, *A Source Book of Gestalt Psychology*. Routledge and Kegan Paul, London, pages 71–88.
- Yannakakis, Georgios N., Roddy Cowie, and Carlos Busso. 2017. The ordinal nature of emotions. In *Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 248–255, San Antonio, TX, USA, October. IEEE.

Appendix

Comparison Table 1

A comparison of anchors of the Drivenness continuum with the ones of the Sensibility continuum. Categorisation is used only to simplify comparisons and verify correspondences between concepts in different fields.

Sensibility	Drivenness	Analysis
Opinions	Endogenous goals	Endogenous goals are linked to stable, personal stances, described by the concepts of sentiments and opinions
Sentiments		
Emotions	Exogenous goals	Exogenous goals are attended on a less personal level than endogenous ones, so they are more linked to emotions as action triggers. They have a relationship, however, with sentiments in that they must be accepted on a personal level before being attended
Feelings	Pseudo-goals	Pseudo-goals are linked to automatic sensible experiences that are closer to perception and reflex reactions
Affect		

Comparison Table 2

A comparison of anchors of the Causality continuum with the ones of the previous continua.

Sensibility	Drivenness	Causality	Analysis
Opinions	Endogenous goals	Imagining	Activities aimed at considering potential outcomes of alternative choices in the past are performed to improve personal performance in the future, building a consolidated opinion system. In this sense, imagining activities are directed towards endogenous goals related, for example, to performance in decision making
Sentiments		Doing	Intervention abilities are aimed at modifying the context of the communicative exchange coherently with goals proper, either exogenous or endogenous. In this sense, they cover both emotions as triggers of decision making processes and sentiments as structured stances towards elements that are relevant for endogenous goals
Emotions	Exogenous goals		
Feelings	Pseudo-goals	Perceiving	Perceiving the context of the communication exchange and reacting to it in a non-conscious way is related to pseudo-goals management and, consequently, to non-conscious sensible experiences
Affect			

A comparison of anchors of the Uncertainty continuum with the ones of the previous continua.

Affect		Feelings	Emotions	Sentiments	Opinions	Sensitivity
Pseudo-goals		Perceiving	Exogenous goals	Endogenous goals		Drivenness
Associative learning			Doing	Imagining	Causality	Uncertainty
			Logical reasoning	Bayesian reasoning		Analysis
			Logical reasoning is the fundamental way to approach goal-directed reasoning, thus acting in an interventionist way to <i>do</i> things with language. Performing reasoning under assumptions, it allows attending externally provided communicative goals as well as planning how to reason about the communication exchange to pursue endogenous goals	Bayesian reasoning is effective to approach goal-directed reasoning with a focus on endogenous goals. Implementing reasoning under uncertainty, it can compute maximum expected utility actions among possible ones and simulate alternative scenarios either retrospectively or prospectively. At higher levels, working on hypothetical scenarios that are not explored by contingency but rather by reflection, it is driven by the less emotionally characterised sensible experiences, sentiments and opinions		

Comparison Table 4

A comparison of anchors of the Initiative continuum with the ones of the previous continua.

Affect		Feelings	Emotions	Sentiments	Opinions	Sensibility	
Pseudo-goals		Perceiving	Exogenous goals	Endogenous goals		Drivenness	
Associative learning			Logical reasoning	Doing	Bayesian reasoning	Imagining	Causality
Backchannel				CRs			QA
		Acks	Analysis				
Backchannel			CRs	QA	Proactivity	Self-access	
			Self-access constitutes the most introspective form of communicative activity, where interaction turns inward and becomes self-referential. At this level, cognition engages Bayesian reasoning to imagine alternative worlds, test counterfactuals, and build moral systems. This reflective mode aims at improving performance in decision making to better attend endogenous goals, commit to future behaviours and analyse produced ones, either for self-understanding or for explanation purposes				
			Proactivity reflects a stage where the agent pursues its own, endogenous goals. By taking the initiative to move the dialogue forward, it deals with a significant amount of uncertainty that is not implied by the more <i>mechanical</i> exchanges at the lower level. It involves reasoning about potential outcomes using probabilistic expectations and structured sensible information for action evaluation purposes. It also relies on logical filtering of impossible context configurations to reduce computational complexity				
			QA exemplify intentional, goal-directed communication that relies on logical reasoning and inferential understanding. QA exchanges pursue explicit problem-solving and knowledge exchange to attend exogenous goals expressed by the interlocutor to obtain information				
			CRs signal a breakdown in mutual understanding and an active effort to repair it through logical, goal-directed reasoning. They add control and linguistic planning, marking a transitional point in the continuum, where agents move from automatic affective responsiveness toward intentional communicative reasoning, therefore <i>doing</i> things with language. At this level, exogenous goals related to attending the interlocutor's intention to transmit a message are managed with emotions signalling inconsistencies in the dialogue state and triggering clarification strategies				
		Acks	Acknowledgments confirm reception without requiring deep reasoning. They pursue the pseudo-goals related to maintaining the interpersonal rhythm of communication and provide low-level validation to the interlocutor. They are rapidly managed associatively at perception level, serving as a crucial foundation for the emergence of more complex communicative acts				
			Backchannels represent the most immediate and reactive layer of communicative behaviour. They are spontaneous, automatic responses attending the pseudo-goals related to indicating attention and engagement, like pose and gaze orientation. They are processes characterised by rapid, affect-driven associative mechanisms. Despite their simplicity, backchannels sustain the emotional synchrony and mutual alignment upon which all higher communicative levels depend				

Comparison Table 5

A comparison of anchors of the Performance continuum with the ones of the previous continua.

Affect		Feelings	Emotions	Sentiments		Opinions	Sensitivity
Pseudo-goals		Perceiving	Exogenous goals	Endogenous goals	Doing	Imagining	Drivenness
Associative learning			Logical reasoning	Bayesian reasoning			Uncertainty
Backchannel	Acks						Initiative
Social signals		Performance	CRs	QA	Proactivity	Self-access	Commissives
			Assertives	Declaratives	Directives	Expressives	Commissives
			Assertives	Declaratives	Directives	Expressives	Commissives
		Social signals express affective alignment and attention through embodied cues. They establish the emotional and perceptual synchrony on which higher communicative and reasoning processes are built. They provide crucial, real-time feedback that continuously shapes communication	Assertive aspects of a speech act are based on logical reasoning about the state of the dialogue representation, the common ground, and attend the exogenous goal of transmitting a message. In this, they are dominant in CRs and are quickly triggered by emotions signalling representation problems. It is the most basic way to use language to <i>do</i> something, like repairing different kinds of inconsistencies	Declarative aspects can be added through logical reasoning to attend both endogenous and exogenous goals. They allow using language to modify the common ground, and to establish facts or institutional states so they are dominant in QA but also play a role in proactive behaviour. Their performative force depends on structured sensible experiences, which provide the affective grounding that gives declaratives legitimacy and credibility within collective contexts	Directive aspects operationalise proactive changes in the communicative context by directing other participants to behaviour commitment. As they are not guaranteed to succeed, they need reasoning under uncertainty capabilities provided by Bayesian reasoning. Structured sensible experiences provide information about an act's success chances and its endogenous value	Expressive aspects manifest the affective and evaluative core of communication. They connect to endogenous goals of alignment, regulation, and self-expression. They can only be added when self-reflection capabilities are present. Through recall, they inform, either internally or externally, about behaviour motivations, supporting both self-monitoring and interpersonal coordination	Commissive aspects express future-oriented intentions and commitments. They imply self-reflection capabilities implemented through Bayesian reasoning, projecting actions beyond the present and integrating prediction, affective motivation, and endogenous goal pursuit. Dealing with commitment towards future behaviour, they are strongly linked to opinions in terms of moral obligations to comply