

Theoretical Construction of an Emotion-Linked Behavior Formula and Its Application in Humanoid Robotics

DongCai Zhao

Key Laboratory of Green Fabrication and Surface Technology of Advanced Metal Materials, Ministry of Education, Anhui University of Technology, Maanshan, 243002, China

Abstract: At present, there exist numerous theoretical formulations regarding emotions. However, once a behavior initiated by a certain emotion is put into practice, the execution of such behavior will introduce additional stimuli, which in turn generate new emotions. In this paper, these influencing factors are comprehensively taken into account, and a behavioral formula is constructed on the basis of drawing on Simonov's emotion theory. This behavioral formula mainly consists of a main-process term and a sub-process term. The sub-process term is triggered by the main-process term and exerts either a promoting or inhibiting effect on the latter. The formula is then illustrated with common examples, including decision-makers who do not participate in behavioral output, individuals who claim a share of the benefits but refuse to engage in any affairs, independent entrepreneurs, wage earners, group members who abandon themselves to degeneration, and those who exert their utmost efforts for the group. This illustration also serves as a verification of the formula's validity. Meanwhile, combined with generative large language models (LLMs) and the Cognitive Process and Information Processing Model (CPIPM), the information generation method in the formula is elaborated, aiming to provide valuable insights for the development of humanoid robots.

Keywords: Humanoid Robotics; Behavioral formula; LLMs; CPIPM; Emotion-Behavior Interaction; Simonov's Emotion Theory

1 introduction

The knowledge base is the core of an expert system [1-2], storing the knowledge of experts. This knowledge includes concepts, the relationships between concepts, and the attributes of concepts. The explanation of a concept is also achieved through these relationships and attributes. Among these relationships, there is a key one: when the subject is a person, this information relates to the emotion-driven attributes of that person. The creation of these attribute values is based on commonly accepted knowledge, thus reflecting a socially accepted average value that is considered correct at a specific time. For instance, when viewing a painting, the emotional value it evokes in an individual is judged according to society's recognition of that painting. Therefore, works by renowned artists or world-class masterpieces are associated with higher emotional value attributes. Clearly, this association may lose its semantic truth for a particular individual. For that individual, if they had previously owned a better masterpiece, the emotional information that this painting can provide may not hold the positive significance suggested by its semantic association. Conversely, if this painting has been a long-term pursuit for the individual and they have paid a great price to acquire it, then the emotional information it provides will carry more positive significance than what its semantic association suggests. In many other cases, the painting may evoke little to no emotional response in the individual, as they may not appreciate it. In such instances, the emotional information related to its semantics may not even arise in their mind, and they might instead focus more on estimating its economic value.

E-mail addresses: dczhao@ahut.edu.cn

The above example illustrates that the construction of an expert system attempts to establish semantic associations of "correct knowledge" within human knowledge. Therefore, these correct semantic associations may appear lifeless from an individual's perspective; while they demonstrate extensive knowledge, their emotional responses are very "average," resulting in an inevitably rigid image.

In the field of artificial intelligence, there is an aspiration to create intelligent robots (embodied AI). If robots are made to express emotions strictly in accordance with the generally recognized emotions represented by each word, they will inevitably come across as very rigid. If associations are established between words and scenarios, the emotional expressions exhibited will still be a kind of averaged state. To enable robots to express emotions in a more personalized manner, this approach draws on emotion theories and combines them with computer generative models to construct emotional associations among pieces of information. If this idea can fortunately be realized, the emotions displayed by robots are bound to be extremely fascinating.

2 The Selection of Emotion-Driven Theory and the Selection of Language Generation Models

When an individual confronts a certain matter, they may experience emotional arousal. When this emotional arousal is translated into action, since new elements are introduced, the emotion may change, and the behavior may also change accordingly. There are numerous emotional models, with some well-known ones including M. R. Arnold's appraisal-excitation theory[3], S. Schachter and J. E. Singer's attribution theory of emotion[4], Richard S. Lazarus's cognitive appraisal theory[5], Pavel Vasilievich Siminov's emotional cognitive theory[6], and emotional dissonance theory[7], et al.[8] These theories have now become common psychological knowledge, with distinct features that make them not universally interchangeable. If one attempts to derive a method for calculating emotions from these theories, build it upon the foundation of artificial language, accomplish the measurement of an agent's emotions, and calculate the behavioral motivations underlying those emotions, it can provide a computational framework for constructing the behavioral patterns of the agent, which refers to an artificial robot with human-like intelligence.

In the emotional models, Siminov's emotional theory is expressed as emotion (E) being equal to the product of need (N) and the difference between necessary information (I_n) and available information (I_a), that is: $E = -N (I_n - I_a)$. In this formulation, no explicit definitions are provided for I_n and I_a ; however, from an intuitive perspective, this expression is highly convincing. Later in his career, Siminov focused on research at the intersection of brain structure and information [9-10], offering an experimental interpretation of information.

If we consider I_n and I_a as forms of linguistic expression, then when an intelligent agent is capable of autonomously generating language, its emotions can be calculated using this method. In terms of language generation by intelligent agents, Large Language Models (LLMs), based on probabilistic principles, are able to produce language that appears highly realistic in most cases, though occasionally they may generate nonsensical or "bullshit" outputs [11-12] with great confidence [13]. This phenomenon is actually a common flaw in generative models; similarly, image

generation models can also produce unfounded yet seemingly plausible images at first glance. The fundamental reason for these shortcomings in LLMs lies in their probabilistic foundation. Therefore, there is a need for a compensatory method for language generation that can be adapted through human feedback at the human-computer interaction interface. Clearly, this remedial measure aims to correct the "bullshit" phenomenon without altering the fundamental principles of the model.

Zhao proposed a cognitive model (CPIPIM) based on psychological principles[14]. Although algorithmically grounded in image recognition technology, this model incorporates the interaction between language and the environment during training, thereby establishing a connection between the trained language and its contextual surroundings. Consequently, when language is generated through LLMs, if the interaction between this language and the environment is verified, it becomes possible to fundamentally overcome the "bullshit" phenomenon and enable the intelligent agent to maintain its opinions if it believes them to be correct. Therefore, given the preliminary conditions for language generation by intelligent agents based on linguistic computation, it becomes feasible to calculate an agent's emotions using language-based methods.

If we assume that the intelligent agent has already acquired language capabilities, then when calculating its emotions, we still need to precisely define I_n and I_a . I_a is actually quite clear—it represents the linguistic expression of the agent's current state. I_n , on the other hand, denotes the agent's expected information. However, the question arises as to how this expected information is formed, which is the content that Section 3 of this paper aims to elucidate. Section 4 to 5 of this paper constructs a behavioral drive formula based on Siminov's emotional theory and concludes with an introduction to the formula through typical case studies[15].

In summary, the chosen emotional theory is Simonov's theory of emotion, and for language generation, we selected the language generated by LLMs corrected with CPIPIM. CPIPIM essentially possesses embodied characteristics; once corrected by CPIPIM, it filters language to achieve personalization and authenticity, establishing a connection between language and the environment. At this point, according to Simonov's emotional theory, it is essential to clarify the contents of I_n and I_a . I_a refers to the current state of being, characterized by phenomenological properties, and in terms of semantic construction, it requires the storage of information as it currently exists. The remaining issue is the generation of I_n .

3 Generation of I_n Based on the CPIPIM

In the introduction to the Cognitive Process and Information Processing Model (CPIPIM)[14], to establish a primal drive for the agent, a most fundamental type of need was defined, referred to as basic willpower and habitual willpower.

The activities of an agent revolve around the satisfaction of its own willpower, and all information is imprinted with the mark of cooperation with humans. In the course of its operation in real-world scenarios, the agent learns through training which objects are essential for satisfying its basic willpower, which rules must be abided by (including causal rules and habitual will), and what kind of behaviors can be adopted to satisfy both the basic willpower and the observance of

relevant rules simultaneously. These behavioral combinations constitute the entirety of the agent's willpower-driven activities. These behavioral patterns become ingrained through collaborative training with humans, essentially forming a composite of decomposed and recombined characteristics from others' behaviors and outcomes - thereby creating an entirely new "other" standard. Since individual cognition emerges from lived experience, the aspects related to mapping and causal cognition inevitably incorporate significant traces of others, establishing this "other" standard that the agent subsequently emulates in its functioning. In other words, during development, the agent internalizes through mapping processes what others attend to as what it should likewise attend to. Supported by essential causal relationships and feedback mechanisms, it participates in social collaboration, gradually forming this "other" standard and continually striving to embody the image demanded by this standard (In the CIPIM framework, volition itself is never explicitly defined as an independent entity. Rather, needs only become determinate when objects are perceived and detected - it is through the appropriation of these objects in socially learned ways that volitional demands are satisfied. Consequently, while identifying volitional needs, the agent cannot objectively apprehend volition per se, but can only become indirectly aware of it through its object-mediated manifestations). The quotation marks around "other" indicate this represents not any specific individual but rather an intentional construct - one that naturally emerges during symbolic combination and incorporates features from multiple others. The establishment of this "other" standard draws from social comparison theory [16-18], while its discrepancy calculations clearly reference cognitive-information theory [6].

In programming implementation, once an acquisition target (I_n) is established, the discrepancy between this target and the current "as-is" information (I_a) constitutes the content that needs to be achieved through symbolic processing. When this symbolic mediation is accomplished, it provides the basis for generating corresponding action outputs, thereby completing the transformation from symbols to actions.

The formation of the "other" standard here emerges as a product of mapping processes and causal logic, establishing criteria on a case-by-case basis. For any single matter during a given period, this standard is not fixed but rather exists as a range - specifically, a distribution of situations that the individual aspires to attain. This aspirational standard inherently carries value judgments, thereby creating benchmarks of "good" and "bad." The individual's target standard represents an achievable level requiring their full effort (The fundamental causal logic imposes this as an essential requirement - without which the target becomes meaningless), around which exists an acceptable range. However, engaging with this range inevitably involves emotional responses, whether joyful or frustrating. Crucially, this "just-within-reach" standard remains influenced by present circumstances, undergoing minor fluctuations through logical computations. Fundamentally, the aspirational standard is determined by what natural language formulations can conceptually bridge. When the target proves too ambitious - when natural language combinations cannot span the gap between the current "as-is" state and the desired "to-be" state - regulatory mechanisms will suppress

the target, lowering it accordingly. The inverse also holds true. However, we must acknowledge that during natural language construction, there exists a persistent tendency to underestimate challenges, frequently leading to overambitious target-setting and consequent frustration (It is important to note that what we refer to as "laws" or "rules" here are actually barriers or tools that the intelligent agent constructs for itself through interactions with the environment, following the principles of mapping and causality).

The standards for what an individual aspires to be are established on a case-by-case basis, encompassing all aspects of personal life, thus forming a convergence of multiplicities. For any single matter, the standard exists as a range of values, constantly fluctuating under the influence of life circumstances that create ripples in the logical flow. This inherent multiplicity and variability make it fundamentally difficult for individuals to articulate their own standards concretely. Expressing "what I aspire toward" naturally becomes problematic—it requires extensive explanation, yet such lengthy exposition transforms comprehensible matters into complex logical labyrinths, drifting further away from concreteness. The challenge of self-knowledge is, therefore, innate. Analyzing this through the developmental stages of logic reveals that logic excels at handling sequential or phenomenally similar matters, through which it intuits essences. However, these essences never fully detach from the particularities of individual cases, and the essences themselves offer no further insight. Among existing theories—such as entropy increase or conservation laws—the Formulas themselves provide little comprehensible meaning. Only when applied to concrete cases do they reveal their power. For humans, matters themselves are aggregates of multiplicities, making it impossible to grasp a single, concrete essence. Instead, we can only approach from different angles, forming what we call "understanding of a person"—a far cruder construct compared to insights into the essence behind natural phenomena. When the "other" in an individual's mind becomes clear and simple, or when the "other" tends to disappear, only then can the individual truly recognize themselves.

In any case, the design of an intelligent agent must incorporate a database of "other" standards—potentially even multiple states—where the stored data may appear chaotic. Yet, this data and its interconnections will form an acquisition target, which in turn becomes the agent's own striving goal.

At this point, a natural question arises: *Could this "other" standard database ever become an empty set?* This seems particularly relevant when considering philosophical traditions like Cynicism or Buddhist asceticism, both of which claim to renounce all pursuits and embrace the void ("all is emptiness").

The CIPM model is bio-inspired, and the aforementioned agent design similarly follows biological principles. Thus, just as Cynicism exists among humans, analogous agents should logically exist in artificial systems. In the agent's behavioral drive, actions are determined by the discrepancy between acquisition targets and the current "as-is" state. If this discrepancy triggers an unpleasurable emotional setting that the agent wishes to eliminate, during the essential causality

phase, the agent may reflect on this value—potentially nullifying its existence entirely. This would result in the disappearance of targets, leading to a state of non-striving. For Cynic philosophers or Buddhist practitioners, this phenomenon clearly doesn't apply, as they either embrace such states or are driven by bodhisattvic vows to save all beings. However, the challenge here mirrors issues faced by humans—the same problems would emerge for agents, necessitating corrective algorithms. This extends beyond the scope of constructing such agents as discussed in this paper and will not be further elaborated here.

4 Introduction and Presentation of Behavior Formula

4.1 The behavior-driving factor

The behavior-driving factor is defined as:

$$B = I_N - I_A \quad (1)$$

The driving force of this factor will cause its value to decrease.

I_N is the Necessary Information value of the behavioral goal, which may be generated either due to emotional drives or additional goals due to external reasons. I_A is the Available Information value corresponding to the behavioral goal. The difference between I_N and I_A creates a driving force that motivates individuals to use behavior to reduce the gap between them. In the explanation of emotions, I_n refers to necessary Information and I_a refers to available Information. They represent different manifestations of the same attribute in different contexts. The information here already includes the demand value N.

In the subsequent calculations concerning behavior, it is necessary to represent the behavior-driving factors at different time instants. Let B_0 represent the behavior-driving factor at the initial moment, and B_t represent the behavior-driving factor at time t .

4.2 Formula for the Manifestation of Purposeful Behavioral Characteristics

Every object exists within the continuum of time, so it inevitably has its own position in time. For an individual, there are stages of growth, work, and retirement, and each stage presents specific tasks. Therefore, the main tasks of their work are determined, and logic can only seek limited changes from the specific circumstances they find themselves in. Thus, for any individual, when determining the calculation for behavioral goals, it is crucial to place them within a goal they are currently striving for. If the self-set goals seem meaningless to others but are matters that require effort to change for the individual themselves, then there is no difference for the individual between establishing recognized goals and establishing goals that only they see as meaningful.

The determination of Behavioral Aim in individuals is described by Formula (2).

$$A_B = r v B_0 \quad (2)$$

In Formula (2), r represents the rationalization factor. Individuals do not evaluate all possible goals as targets. When evaluating a possible goal, it either should not disrupt the current main behavioral process or should involve significant factors that can evoke strong emotions for the individual during the execution of the current main behavioral goal. Even if the above conditions are met, if the existence of a gap is considered to be causally related or is not a requirement that the

current tasks need to address, it does not constitute a new behavioral process goal, and the value of r is set to 0. And v represents the self-value factor. For individuals with self-confidence, this value is 1. v is a value highly related to the event. The introduced v factor is based on the big-five theory of personality. The research on the Big Five personality traits is developing towards a more detailed direction[19-21], and this paper will not delve into its specific content. The Big Five personality traits consist of five parameters, with three of them used in the behavioral formula. Neuroticism corresponds to self-value (v), conscientiousness corresponds to perseverance (p), and openness corresponds to wise (w). Extraversion and agreeableness are manifestations of behavioral output, and they are not included in the internal model of behavioral evaluation here. If it is necessary to carry out calculations in the behavioral formula, the values of these three parameters must be controlled within the range of 0 to 1. This characteristic is determined by the design features of the formula. The coefficients introduced here are assigned meaning through the explanation of the Big Five personality traits. From the perspective of mathematical correction, they are constant terms that need to be determined experimentally. This is one of the most commonly used methods in mathematical fitting.

In behavioral goals, it is possible to operate in a multi-threaded manner, but the main thread must be the expression related to learning, work, or other aspects within the current environment. Other threads must operate while ensuring the integrity of the main thread. Once actions start to be executed, they become an output sequence that operates according to the formula for the manifestation of purposeful behavioral characteristics (Formula (3)). Behavioral goals may be multi-threaded, but the operation of the logical stream cannot concurrently pursue multiple goals. It can only be executed one by one, allowing for switching between multiple goals. Actions in behavioral output must be singular but can possess a dual meaning. The concept of logical stream originates from the stream of consciousness, with the term "logical" implying a computational aspect. The concept of the stream of consciousness has been widely discussed by associationist philosophers such as David Hume and introspectionist psychologists such as William James, and it remains a topic of research today [22-27].

Once the aim is determined, it is time for behavior action output, as indicated in Formula (3).

$$B_Process=r[pf(t)B_0 + c_RG \cdot B_0 \cdot (B_0 - B_t)/B_0 + T_D - (B_B + wB_{B1}) - (B_{ot} + wB_{ot1})] \\ =r[pf(t)B_0 + c_RG \cdot (B_0 - B_t) + T_D - (B_B + wB_{B1}) - (B_{ot} + wB_{ot1})] \quad (3)$$

$pf(t)$ represents the persistence factor, which refers to the level of behavioral excitement derived from the desire for a goal. The function $f(t)$ is a law of memory decay, which refers to the phenomenon that a target is easily forgotten if it is not repeatedly activated, as revealed by Hermann Ebbinghaus. Since the discovery and publication of the laws of memory, this field of research has continuously attracted the attention of scholars [28-29]. Here, we do not intend to conduct a detailed analysis of its specific patterns and numerical values, but rather introduce it as an influencing factor. p is the persistence coefficient, expressing the degree of individual's dedication to a behavior, w represents the wise factor, which expresses an individual's level of rationality. The introduced p and

w factor is based on the big-five theory of personality.

The expression $c_RG \cdot B_0(B_0 - B_t)/B_0$ represents the achievement factor, which is an individual's own judgment of the value of their work. On the other hand, c_RG stands for the social recognition coefficient, referring to an individual's perception of whether their work is recognized within society. The introduction of c_RG is based on the theory of need for achievement proposed by the American behavioral scientist David McClelland. Recent research on the need for achievement has focused on occupations and interpersonal relationships [30]. In this paper, it is merely incorporated as a factor in the formula. Regarding specific numerical values, further detailed research is required for supplementation. Nevertheless, it is certain that this represents a highly personalized value.

The B_B refers to the emotions directly triggered by actions driven by B_0 , excluding the information already covered in the calculation of B_0 . B_{B1} is the emotion value generated based on the extension of potential events triggered by B_B . In general, it represents the inhibition of B_B , which is related to moral and social customs and is not included in the calculation of B_0 .

T_D represents the factor for the desire for triumph, an individualized factor that arises along with the emergence of behavior. When individuals are placed within society, they manifest the need for achievement. However, inherently, individuals also possess the motivation to overcome obstacles and survive, which is a trait similar to animals. The specific numerical values related to this trait are not discussed here.

B_{ot} refers to the emotional value unrelated to the specific matter, i.e., emotions triggered by information that is not directly related to the calculation of B_0 , B_B , and B_{B1} . B_{ot1} represents the emotion generated based on the extension of potential events triggered by B_{ot} . Generally, it represents the inhibition of B_{ot} , which is related to moral and social customs.

The terms in Formula (3), $pf(t)B_0 + c_RG \cdot (B_0 - B_t) + T_D$, correspond to the main-process items (B_Process_main). The middle two terms, $-(B_B + wB_{B1})$, represent the sub-process items (B_Process_sub). The last two terms, $-(B_{ot} + wB_{ot1})$, represent the disturbance items (B_Process_dis). The main-process items consist of three components. The first component is determined by established goals and is often the prerequisite for an individual's survival. Therefore, it must be executed and ensured. The following two components serve to reinforce the main-process. The sub-process items are generated by the main-process and do not hinder its completion. However, they may overshadow the main-process. The disturbance items represent emotions triggered by matters unrelated to the process. They are usually temporary and do not have a significant impact on the entire process. Therefore, Formula (3) can be simplified to Formula (4). Of course, everyone encounters situations with disturbance items, and sometimes it may be necessary to interrupt the main-process and handle the disturbance items. If the disturbance items are so strong that an individual cannot persist with the main-process, the matter at hand will disappear, and the content expressed in Formula (3) will need to be updated. When the sub-process item is greater than the main-process item, the matter on which the main-process item relies itself loses the significance of persisting.

$$B_Process = r[pf(t)B_0 + c_RG \cdot (B_0 - B_t) + T_D - (B_B + wB_{B1})] \quad (4)$$

The completion of B_0 relies on a certain event, and the calculation is based on the available information after the completion of the event compared to the available information without the event. This often creates a strong driving force. The actions and events triggered by the completion of the event stimulate B_B . The completion signs of B_0 based events and the purpose of B_0 based events may sometimes be inconsistent, which will provide opportunities for B_B to put into practice and encourage the output actions to move closer to the actions required for the completion sign. When $B_B \leq 0$, it is an ideal situation. When $B_B > 0$, it will cause interference to B_0 , and then wB_{B1} is needed to suppress B_B . When the sub-process item is not beneficial to the main-process item, the sub-process item will immediately prompt the individual to generate behavioral goals under the main-process item, and it may lead to the establishment of a new sub-process. When achieving the completion sign, if the intervention from others causes B_B to become a negative value, it will facilitate the individual in accomplishing their own goals. However, if the intervention from others causes B_B to become a positive value, the number of sub-processes established based on the sub-process item may become extremely large and variable. Moreover, these sub-processes, due to their own logical constructions, may go beyond the scope of the matters related to B_0 , resulting in often dire consequences, commonly known as "internal collapse" in the usual sense. In the main-process item, due to the effect of attenuation, the goal may gradually be lost, leading to the loss of that process. Therefore, if one wants to persist with the goal, it is necessary to evaluate the goal in a timely manner, establish it again, and engage in self-motivation frequently.

The sub-process item exists in relation to the main-process item. However, in terms of individual behavior, the main-process item sets the overall goal. When progressing towards this goal, emotions regarding the goal tend to be calm. In other words, individuals do not engage in excessive thoughts about the goal as they move forward under the established goal. This characteristic is determined by the nature of logical stream (single-threaded). At the same time, the nature of logic stream determines that emotions arising from current stimuli become the primary focus of individual experiences. Therefore, the sub-process item tends to stay longer in the mind than the main-process item.

5 Examples of a behavioral formula

5.1 Case 1: Decision-maker who does not participate in behavioral output

In some cases, there is no need to make a decision to actively output behavior, and the output can be achieved simply by speaking, for example, by using one's voice to exert powerful influence. In such situations, formula (4) will be transformed into formula (5) or formula (6).

$$B_Process = r[pB_0f(t) + c_RG \cdot (B_0 - B_t) + T_D] \quad (5)$$

Or

$$B_Process = r(B_B + wB_{B1}) \quad (6)$$

Formula (5) represents an expression that ignores the sub-process term and only includes the main-process term. It is an ideal form of representation for Formula (4), but in reality, it is difficult

to fully satisfy. However, when the executor does not need to output any actions directly related to the dependent matter B_0 , Formula (5) naturally becomes the expression of Formula (4). In the output of matters, when the person involved in the output is only superficially involved and does not need to make any direct actions related to the matter, they become absurd participants. Moreover, they are participants who claim to be on the side of "righteousness," and thus, always appear to be correct. However, their involvement is often harmful to the completion of the matter dependent on B_0 . As the matter is pushed forward, those who participate according to Formula (5) will automatically supplement the content related to the sub-process term, which often contradicts the content of the sub-process term formed in the minds of the individuals actively pushing the matter forward. This content often carries aggressiveness (hostility) and may lead the matter into a deadlock of internal conflicts.

Formula (5) describes a scenario where, despite not participating in the action, there is still a driving force for the completion of the matter that B_0 relies on. The situation outlined in (6) involves neither a driving force for participation nor any actual form of participatory output; it represents the evaluator's perspective. Please refer to the following case for further illustration.

5.2 Case 2: Persons who substitutes themselves into the distribution of benefits but does not participate in any tasks

The B_B in Formulas is triggered by actions related to B_0 , and B_{B1} is triggered based on possible actions of B_B , when B_0 does not exist, B_B and B_{B1} also do not exist. Therefore, it can be seen that the sub-process term is a companion term of the main-process, existing alongside the main-process and disappearing with the disappearance of the main-process.

In reality, due to humans' inherent causal-logical nature, individuals tend to place themselves in a given context mentally, yet without the need to participate in its execution. Consequently, they take the achieved outcomes for granted. As a result, in terms of goal execution, there is no driving force—that is, $B_Process_main$ disappears. However, the goals achieved by the main process items will evoke a functional element in their minds, similar in form to $B_Process_sub$ (see Formula 6). Specifically, among those who do not participate, Formula (4) should not logically exist, yet sub-process items are retained without preserving the driving items. When such items persist, it becomes evident that individuals who are not supposed to be involved are, in fact, mentally engaged, or rather, some have mentally placed themselves in the scenario. These self-inserted participants clearly do not contribute labor or partake in benefit distribution, making their calculated outcomes entirely inequivalent to those of actual participants. Therefore, Formula (6) cannot be classified as a sub-process item; instead, it should be regarded as an addendum to the main-process items, representing the perspective of the evaluator. Being an addendum, it does not include a negative sign.

Whether decision-makers or evaluators, when they take actions based on their own emotions, they will each generate their respective sub-process items. Meanwhile, considering that the actual executors of the tasks will also produce sub-process items, it becomes evident that the sub-process items generated by these three groups are not equivalent and lack comparability. If these three

groups are composed of different individuals, it can be anticipated that the sub-process items will also be inequivalent among individuals, although there may be a higher degree of similarity across different groups.

5.3 Case 3: Entrepreneurs

When an individual engages in autonomous entrepreneurship, the self-established goal driving force is referred to as B_0 . Throughout the entire process of action, any changes caused by behavioral stimuli will lead to changes in B_0 , which means that the entrepreneurial goals can be altered due to factors generated by other emotions. The emotions of the sub-process item still exist, but they do not pose a threat to the B_0 -based item, so they were cancelled.

The aforementioned process is a very linear one, where (4) is transformed into (5). However, in reality, as individuals progress towards their goals, they inevitably face various obstacles. Overcoming these obstacles and completing the tasks associated with B_0 will often trigger emotions unrelated to the specific tasks involved. These emotions may guide to take actions specifically targeted at addressing these emotional aspects, such as eliminating certain hindering factors. If the elimination of these hindrances is solely for the purpose of completing B_0 , it does not constitute a separate sub-process but is encompassed within B_0 . However, if the elimination of hindrances extends beyond the completion of B_0 , meaning it goes beyond the scope of entrepreneurship and establishes new goals, it does not transform (5) into (4). The new process does not harm the main-process; it simply adds an additional process alongside it.

5.4 Case 4: Wage Workers

Wage workers rely on B_0 as the basic guarantee for satisfying their various needs. Therefore, for the tasks underlying the completion of B_0 , they must be fulfilled as long as they are part of the work position. However, compared to the original intention of setting up these tasks in the work position, wage workers, although they are consciously aware of the purpose of their position, generate a stimulation value, $-(B_B + wB_{B1})$, during the process of completing that purpose, while meeting the completion sign for B_0 . If the supervisor of the wage worker can decompose the purpose of the underlying tasks to the extent that the completion sign and the purpose of the tasks perfectly align, then for the wage worker, the behavioral driving formula as (5). However, if the decomposition cannot reach this level, then $-(B_B + wB_{B1})$ will inevitably exist, and it will prompt the individual to engage in tasks that do not correspond to the original intention while meeting the completion sign. Of course, in some cases, there may be a phenomenon where $-(B_B + wB_{B1})$ strengthens the main-process.

It is evident that wage workers, driven by $-(B_B + wB_{B1})$, may engage in behaviors that harm the tasks underlying B_0 , causing losses to those directly involved in the event's benefits. Therefore, within the organization, there must be a set of methods to address this situation of covert non-compliance. The most effective approach in these methods is to minimize the value of $B_B + wB_{B1}$. This approach has been widely promoted in production line operations, but treating humans as machines is not an ideal practice. From a technological standpoint, it should be the goal to replace

positions with machines. However, for many management and research positions, tasks cannot be decomposed to achieve a value of $-(B_B + wB_{B1})$ equal to zero. Thus, trying to adopt management methods based on a value of $-(B_B + wB_{B1})$ equal to zero in these cases would be contrary to the facts and would often generate more $-(B_B + wB_{B1})$ items or increase the value of $B_B + wB_{B1}$, resulting in more harm than benefit. In this situation, the best approach is to clearly communicate the objectives, promote self-actualization, and provide incentives based on outcomes, thereby motivating participants to willingly and consciously complete the tasks underlying B_0 . The generation of $-(B_B + wB_{B1})$ occurs not only due to the difference between task objectives and completion sign but also extends beyond these factors. The previous explanation focused on minimizing the value of $B_B + wB_{B1}$ resulting from the difference between task objectives and completion sign. Other factors that contribute to its generation often stem from comparisons among individuals or the meaning of work, ultimately raising questions of fairness and justice.

More directly speaking, it means that the available information and necessary information upon which B_0 relies tend to become equal. Even for individuals who think actively, since the matters they contemplate are highly likely to exceed the scope of those upon which B_0 relies, the probability of their thoughts exerting a driving force on the matters related to B_0 is also very low.

5.5 Case 5: Individual Self-indulgence within a Group

In this case, the term "group" has two implications. Firstly, it refers to a large population where most members are unfamiliar with each other, meaning that they do not exchange thoughts and emotions. Secondly, it indicates that the behavior of all individuals within the group is predetermined, and they do not need to take responsibility for their own future, but only for the assigned tasks.

Individuals within a group are still driven by B_0 , but in a structured group, due to the prescribed nature of information, the necessary and available information for individuals is prescribed, leading to a tendency for the two to converge in an individual's perception. In this case, emotions are stable. The main-processes do not truly drive the individual; as long as the individual participates in the main-processes, they receive the rewards defined by those processes. On the other hand, while an individual's work within a group may indeed alter the available information, this change is minimal and has little impact on the actual available information for oneself. Therefore, the driving force of B_0 -reliant tasks on individuals is very weak.

Although emotions may be stable in this situation, it is not necessarily the case that the information obtained from the event is equivalent to what the individual's desires. If the Available Information is superior to the Necessary Information, it adds a sense of joy to the stable emotions. Conversely, if it is inferior, it adds a sense of pain to the stable social emotions. This means that the sub-process can either promote or hinder the main-process, or even override it.

If the addition of joy or pain to the stable emotions does not lead individuals to reflect on the event corresponding to those emotions, it implies that individuals are simply going with the flow. Going with the flow signifies a lack of responsibility towards the event associated with stable

emotions. If the reflections of some individuals are capable of questioning the event corresponding to the stable emotions, the results of that reflection may disrupt the event itself. In fact, reflecting on an event is incredibly challenging. The event involves numerous complex factors that cannot be easily deduced through smooth and natural reasoning. Consequently, the conclusions drawn through reasoning can be numerous, but often remain unverifiable, leading individuals to rely on one theory. When individuals invest in one or a few routes, it signifies going with the flow, while neglecting other possibilities.

5.6 Case 6: Individual Wholehearted Effort within a Group

This case is a continuation of the previous case, as astute readers may have already identified a loophole in the emotional reasoning described earlier: the fact that the I_A and I_N are sometimes unpredictable and cannot be predetermined.

For individuals, in their daily affairs, the I_N is determined through an estimation of their own actions. Thus, individuals take action to align the expected I_A with the I_N . However, in certain situations, the I_N cannot be predicted, and the assessment of I_A can only be based on the current circumstances. The individual's actions can only lead to a potential improvement in I_A , but the extent of improvement cannot be determined. This is often an unfavorable situation where individuals have lost their ability to predict the future. If individuals have developed a Persistence character (p), they may continue to make efforts to change the situation and regain control over their future. Conversely, they may fall into a state of despair and give up. When a group, as defined in case 5, faces such a situation, the premise of stable emotions described in Case 5 no longer exists. The characteristics of their emotions align with those of individuals facing the situation. Therefore, individuals face the choice of making efforts to change the situation or accepting their fate. However, due to the presence of the group, most individuals tend to choose to make efforts to change. Analyzing the Behavior Formula (4), the value of B_0 is significant in this context, and the only way to achieve the desired outcome is to change the state of the matters on which B_0 relies. This is why unity and collective strength are emphasized.

Individuals choose to make efforts to change due to the motivation derived from the efforts of those who strive for change and the inspiration of unified planning. They see hope and actively participate in the change, exerting effort towards making a difference. As mentioned earlier, the I_N and I_A cannot be predicted. Therefore, unified planning can only serve as inspiration. Individuals choose to believe through their self-defense mechanisms and thus take action. Here, the powerful logical capabilities of humans are clearly demonstrated. When unified planning can be predicted, individuals no longer exhibit belief in the future of the plan. In reality, because of belief, their emotions are not stirred by $B_Process_sub$, and there is no driving force to take action. However, when unified planning cannot be predicted, individuals display belief, but due to their lack of actual belief, their emotions are stirred, leading to rapid and effective action. In historical context, it is ironically observed that when unified planning can be predicted, the lack of action from individuals makes the outcome unpredictable, leading to disastrous consequences. However, disasters

themselves fail to awaken people to unite and work together. Conversely, when the result of unified planning cannot be predicted, individual actions contribute to positive outcomes, helping people emerge from the uncertainty of the unpredictable future.

In such situations (typical in wars and disasters), it is evident that individuals within a group can unleash tremendous power, and the strength of collective unity is immense. When people are separated, they lose their mutual dependence, and the breaking point for predicting I_N no longer exists, leaving only suffocation and despair. Therefore, it is very meaningful to form a unified command as a whole.

Of course, in unpredictable situations, some people may fall into pessimism and despair and stop striving.

6 Conclusion

Formalizing human behavior is an important step in aligning human disciplines with the natural sciences. This attempt seeks to summarize human behavior in a formalized manner. If a corresponding database for I_A and I_N can be established, and personalized calibrations can be assigned to the parameters in equation (4), then the significance of equation (4) will become evident. Furthermore, through the classification and summarization of the I_A and I_N databases, explanations can also be provided for concepts such as jealousy, boasting, comparison, and the anchoring effect, which will not be discussed further in this paper.

What is the value of generating the meanings of I_A and I_N through generative algorithms (MMLs and CIPM)? Furthermore, in terms of this discussion, developing algorithms for I_A and I_N is a substantial undertaking. I_A is currently characterized by its phenomenological properties, and the basic theory surrounding this information is a conclusion of phenomenology; I_N draws from social comparison theory. Therefore, the theoretical materials for I_A and I_N are relatively rich. However, in terms of information generation, one can start with the simplest methods. For instance, for an intelligent agent, the information that the agent can fully control can be treated as the dataset for I_A , while the information that it cannot control but frequently encounters can be treated as the dataset for I_N (the theoretical foundation of this dataset remains based on probability theory, which is consistent with the theoretical foundation of generative models). Then, using semantic analysis, one can impose a restriction on the I_N dataset based on whether the information that the agent can control can reach the I_N dataset (this restriction acknowledges the logical reasoning capacity of humans). This would complete the preliminary construction of the dataset. Once these datasets are established, the corresponding emotional data can then be constructed based on the behavioral formulas (4), (5), and (6).

When the datasets for I_A , I_N , and emotional values are fully constructed, it amounts to creating training materials related to emotions, which is the objective of this paper. If this data is extended to embodied intelligence, it means that by constructing behavioral formulas, we can establish goals for the behavior of intelligent agents. This process is quite complex, but if it is confined within a limited scope, it is still feasible to construct algorithms through computational means.

References

- [1] E. E. Ogheneovo, P. A. Nlerum, Knowledge representation in artificial intelligence and expert systems using inference rule, *Int. J. Sci. Eng. Res*, 11(4) (2020) 1886-1900, https://www.researchgate.net/publication/347504395_Knowledge_Representation_in_Artificial_Intelligence_and_Expert_Systems_Using_Inference_Rule
- [2] D. Janjanam, B. Ganesh, L. Manjunatha, Design of an expert system architecture: An overview, In *Journal of Physics: Conference Series* 1767(1) (2021)012036. IOP Publishing. <https://doi.org/10.1088/1742-6596/1767/1/012036>
- [3] M. B. Arnold, *Emotion and personality*, Columbia University Press, 1960.
- [4] S. Chachter, J. Singer, Cognitive, social, and physiological determinants of emotional state, *Psychological Review*, 69(5)(1962)379 – 399. <https://doi.org/10.1037/h0046234>
- [5] R. S. Lazarus, *Emotion and adaptation*, Oxford University Press, 1991
- [6] P. V. Simonov, The need-informational theory of emotions, *International Journal of Psychophysiology* 1(3)(1984) 277-289. [https://doi.org/10.1016/0167-8760\(84\)90047-3](https://doi.org/10.1016/0167-8760(84)90047-3)
- [7] E. Aronson, *Dissonance theory: Progress and problems*, *Theories of cognitive consistency: A sourcebook* 249(253) (1968) 12.
- [8] R. Plutchik, H. Kellerman, Eds, *Theories of emotion* (Vol. 1), Academic press, 2013.
- [9] P. V. Simonov, Need-informational interaction of brain structures, *The Pavlovian journal of biological science: official journal of the Pavlovian* 1(1981)125–130. <https://doi.org/10.1007/BF03003216>
- [10] P. V. Simonov, M. V. Frolov, Informational need of emotional stress, *Acta Astronautica* 8(9–10)(2003) 1043-1049. [https://doi.org/10.1016/0094-5765\(81\)90077-1](https://doi.org/10.1016/0094-5765(81)90077-1)
- [11] N. Jones, AI hallucinations can't be stopped—but these techniques can limit their damage, *Nature* 637(8047) (2025) 778-780. <https://doi.org/10.1038/d41586-025-00068-5>
- [12] H. Farrell, A. Gopnik, C. Shalizi, J. Evans, Large AI models are cultural and social technologies, *Science* 387(6739) (2025) 1153-1156. <https://doi.org/10.1126/science.adt9819>
- [13] A. Zou, L. Phan, S. Chen, J. Campbell, P. Guo, R. Ren, ..., D. Hendrycks, Representation engineering: A top-down approach to ai transparency, *arxiv preprint arxiv* (2023) 2310.01405. <https://doi.org/10.48550/arXiv.2310.01405>
- [14] D. Zhao, Cognitive process and information processing model based on deep learning algorithms, *Neural Networks* 183(2025) 106999. <https://doi.org/10.1016/j.neunet.2024.106999>
- [15] D. C. Zhao, *The Human nature principle of artificial intelligence*, American academic press, 2022.
- [16] J. Suls, L. Wheeler, Social comparison theory, *Handbook of theories of social psychology* 1 (2012) 460-482. <https://doi.org/10.1007/978-1-4615-4237-7>
- [17] L. Festinger, Social comparison theory, *Selective Exposure Theory* 16(401) (1957) 3.
- [18] L. Wheeler, A brief history of social comparison theory. In *Social Comparison* (1991) (pp. 3-21). Routledge.
- [19] Q. F. Wan, L. Jiang, Y. H. Zeng, A big-five personality model-based study of empathy behaviors in clinical nurses, *Nurse Education in Practice* 38(2019)66–71. <https://doi.org/10.1016/j.nepr.2019.06.005>
- [20] P. Jawinski, S. Markett, C. Sander, J. Huang, C. Ulke, U. Hegerl, T. Hensch, The Big Five Personality Traits and Brain Arousal in the Resting State, *Brain Sci.* 11(2021)1272. <https://doi.org/10.3390/brainsci11101272>

- [21] C. Montag, C. Sindermann, D. Lester, K. L. Davis, Linking individual differences in satisfaction with each of Maslow's needs to the Big Five personality traits and Panksepp's primary emotional systems, *Heliyon* 6 (2020)e04325. <https://doi.org/10.1016/j.heliyon.2020.e04325>
- [22] K. Christoff, Z. C. Irving, K. C. Fox, R. N. Spreng, J. R. Andrews-Hanna, Mind-wandering as spontaneous thought: A dynamic framework, *Nature Reviews Neuroscience* 17(2016)718. <https://doi.org/10.1038/nrn.2016.113>
- [23] Z. C. Irving, Mind-wandering is unguided attention: accounting for the “purposeful” wanderer, *Philos Stud* 173(2016) 547–571. <https://doi.org/10.1007/s11098-015-0506-1>
- [24] Z. C. Irving, A. Glasser, Mind-wandering: A philosophical guide, *Philosophy Compass* 15 (2020) e12644. <https://doi.org/10.1111/phc3.12644>
- [25] J. L. Singer, Navigating the stream of consciousness: Research in daydreaming and related inner experience, *American Psychologist* 30(7)(1975)727–738. <https://doi.org/10.1037/h0076928>
- [26] C. Sripada, A. Taxali, Structure in the stream of consciousness: Evidence from a verbalized thought protocol and automated text analytic methods, *Consciousness and Cognition* 85(2020) 103007. <https://doi.org/10.1016/j.concog.2020.103007>
- [27] L. Tonello, M. Cocchi, F. Gabrielli, J. A. Tuszynski, Stream of consciousness: Quantum and biochemical assumptions regarding psychopathology, *Medical Hypotheses* 101 (2017)78–84. <https://doi.org/10.1016/j.mehy.2017.02.013>
- [28] J. T. Wixted, Absolute versus relative forgetting, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 48(12)(2022) 1775–1786. <https://doi.org/10.1037/xlm0001196>
- [29] Sala S. Della, A. Baddeley, N. Peng, R. Logie, Assessing long-term forgetting: A pragmatic approach, *Cortex* 170 (2024) 80-89. <https://doi.org/10.1016/j.cortex.2023.11.009>
- [30] O. Dousin, N. Collins, T. Bartram, P. Stanton, The relationship between work-life balance, the need for achievement, and intention to leave: mixed-method study, *Journal of advanced nursing* 77 (2021) 1478–1489. <https://doi.org/10.1111/jan.14724>