El problema del marco en la cognición, las emociones y la toma de decisiones

The frame problem in cognition, emotions, and decision-making

MARIA INES SILENZI
Instituto de Investigaciones Económicas y Sociales del Sur (IIESS/CONICET),
Departamento de Humanidades
Universidad Nacional del Sur, Argentina

Recibido: 24/03/2025 Aceptado: 08/01/2026
Publicación en avance: DOI: 10.24310/contrastes.31.2.2026.21533

RESUMEN

Nuestra principal contribución se centra en el debate ya presente en el ámbito de la Inteligencia Artificial en torno a los procesos cognitivos que determinan la relevancia en la toma de decisiones, y cuestiona el papel de las emociones en su resolución. Este debate da cuenta de un fenómeno cognitivo central, el problema del marco, que examina cómo los humanos, en la mayoría de los casos, determinan la relevancia eficientemente. El objetivo de este artículo es explorar los límites y el alcance de la función epistémica de la saliencia y la relevancia de las emociones para abordar el aspecto eficiente del problema del marco.

PALABRAS CLAVE
PROBLEMA DEL MARCO; INTELIGENCIA ARTIFICIAL; COGNICIÓN; EMOCIÓN; TOMA DE DECISIONES.

ABSTRACT

This paper contributes to an ongoing debate in Artificial Intelligence on the *cognitive* processes involved in determining relevance in *decision-making*, while reassessing the role of *emotions* in overcoming this problem. This debate addresses a central cognitive phenomenon, the *frame problem*, which concerns how humans typically determine relevance efficiently. The study examines the limits and scope of the epistemic function of emotional salience and relevance in resolving the two main difficulties associated with the *efficiency aspect* of the frame problem.

KEYWORDS
FRAME PROBLEM; ARTIFICIAL INTELLIGENCE; COGNITION; EMOTION, DECISION MAKING.

I-*Introduction*

Based on the general association between two or more sciences, referred to as a *disciplinary relationship* (Rabossi 2002), our article emphasizes the relationship be-

tween Artificial Intelligence (AI) and a newer field research, the Epistemology of Emotions.

The dialogue established within this 'interface zone' has increased notably in recent decades through three main approaches: conceptual development, critical evaluation, and the adoption of shared findings and perspectives. In this regard, this paper examines how different theoretical perspectives on emotions address a central problem in AI, the *frame problem*[1], which concerns the simulation of cognitive processes involved in determining relevance[2]. This article emphasizes the role of emotions in overcoming this problem, as epistemological research on emotions has advanced in areas of growing interest to AI (Deonna & Teroni 2012; Brun & Kuenzle 2016; Candiotto 2019; Evans 2004).

Some researchers, particularly in the field of affective computing[3], highlight the importance of this relationship, as AI "has ignored a crucial component of basic human intelligence for human problem-solving abilities: the use of feelings and intuition to guide reasoning and decision making" (Picard 2000, pp. 221-2). Research has reported progress in modeling emotional behaviors related to facial expression and vocal intonation, as well as behaviors that contribute to successful goal achievement in decision-making contexts (Politou et al. 2017; Assuncao et al. 2022). From an epistemological perspective, this study therefore asks whether artificial agents could be attributed emotions that collaborate in determining relevance as successfully as humans do.

In addition to the relatively few investigations that focus specifically on this interface zone, most existing proposals lack clarity regarding which aspects of the frame problem they address, a limitation that has generated philosophical controversies[4]. Several debates have emerged from omitting a particular relationship between the definitional difficulty (the complexity of defining the problem) and the resolutive difficulty (the complexity of resolving it). Clarification of the specific aspects of the frame problem under consideration is therefore required before proposing a remedy, since a solution for one aspect may not apply to another.

To elucidate this relationship and avoid confusion and disagreement about the limits and scope of emotions in addressing the frame problem, Section 2 examines the *efficient aspect of the frame problem* (EAFP), which concerns how successfully an artificial agent can determine relevance. Section 3 presents epistemological positions on the capacity of the salience function of emotions to address this aspect and analyzes ways to overcome its main difficulties. This section contributes an overview of the capacity of emotions to solve the defined aspect and difficulties, a topic that the specialized literature has not examined. Section 4 proposes different degrees of epistemic success by considering the requirements of promptness and adequacy. In this regard, some epistemological positions satisfy both adequacy and promptness requirements, whereas others satisfy only promptness. Section 5 presents the conclusions.

The delimited interface zone and the proposed connections demonstrate progress toward an interdisciplinary framework. This approach does constitute a closed or definitive solution, but rather a promising avenue where several postulates of the Epistemol-

[1] This problem was first discussed in McCarthy and Hayes (1969) regarding the situation calculus. From a historical perspective, the failure to solve the frame problem has, for some AI researchers, explained the lack of progress in the field, as the problem ultimately proved to be far more complex than initially assumed (Crockett 1994).

[2] For this interpretation of the frame problem, see Glymour (1987) and Majeed (2020), among others.

[3] The term *affective computing*, proposed by Picard (2000, p. 3), "relates to, arises from, and deliberately influences emotion". For discussion of the origins of this field, its early development, and its applications, see, among others, Calvo et al. (2015).

[4] In this regard, see Fodor (1983), Dennett (1984), and Brown (1987), among others.

ogy of Emotions require reconsideration and reformulation to address other key problems in AI.

II. *The definitional difficulty: The efficient aspect of the frame problem*

This section provides a brief overview of the EAFP, which has generated most of the controversy in the literature. Before examining this aspect, clarification of the meaning of AI is required. In this study, AI is understood "as the ability of a system to interpret external data, learn from that data, and use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein 2019, p. 17).

Among the abilities that an artificial agent must simulate in such flexible contexts, this paper focuses on those involved in determining relevance. In line with Minsky (1968, p. 5), the central challenge for AI is "to make machines that do things that require intelligence when those things are done by human beings". Within this range of processes, the scope of artificial agents in determining relevance as humans do remains a central question.

Definition of *relevance* is also necessary, since this concept is polysemic and may generate confusion, while also playing a key role in characterizing the EAFP. Because the EAFP operates within task-oriented contexts, this study adopts the interpretation proposed by Hjørland (2010, p. 229): "something (*x*) is relevant to a task (*T*) if *x* increases the probability of achieving the goal (*O*) that is implied by *T*". Under this definition, the relevance of a given piece of information depends on its contribution to the successful fulfillment of the objectives implicit in the task. This action-oriented understanding is central to the present investigation, as it emphasizes the fulfillment of predefined purposes.

The EAFP concerns how an agent selects relevant information from a vast data set to achieve a given task with a certain degree of success. In the literature, scholars generally accept that relevant information appears in human cognition in an appropriate manner and at the right time, a phenomenon that characterizes the EAFP. This regularity suggests that humans tend to determine relevance successfully in most cases, in terms of both adequacy and promptness, although occasional errors may occur.

Dennett's robots illustrate this phenomenon. In this example, Dennett (1984) emphasized dilemmas that artificial agents face when confronted with the EAFP. The thought experiment involves a robot (*R1*) whose designers assign the task of retrieving a spare battery located inside a car in a locked room, where a bomb is programmed to explode shortly. *R1* infers that the 'Take Out action (*Car, Room*)' will achieve the task. The robot enters the room and removes the car containing the spare battery. Because the bomb is also inside the car, the robot explodes after a few seconds. To prevent such outcomes, the designers of *R1* conclude that robots must account for both the intended and secondary consequences of their actions.

The designers then construct a second robot, *R1D1*, which also locates the battery and devises a plan of action. However, while the robot infers that moving the car out of the room will not alter the color of the room's walls and begins to evaluate the next implication, the bomb explodes. In response, the designers conclude that the robot must learn to differentiate, through inference mechanisms, between implications that are pertinent to action planning and those that are not. This distinction would prevent the robot from not becoming stalled while classifying each implication as relevant or irrelevant.

The designers then build a third robot, *R2D1*, and assign it the same task. After locating the battery, the robot remains outside the room. Concerned by its inaction, the

designers urge the robot to act, to which it replies: 'I am doing it (...) I am diligently ignoring the thousands of implications that I have determined to be irrelevant! As soon as I discover an improper implication, I add it to the list of what to ignore!" While the robot concentrates on excluding unnecessary considerations, the bomb explodes, and the robot exhausts its battery power.

This example shows that solving the EAFP requires an artificial agent to satisfy at least two conditions: *i*) making appropriate decisions (*adequacy requirement*) and *ii)* doing so within a reasonable time frame (*promptness requirement*). None of Dennett's three robots meets both conditions. The first robot lacks a mechanism that enables full consideration of the implications of its actions. The second possesses this capacity but fails to distinguish relevant consequences from irrelevant ones. The third acquires this latter ability but cannot coordinate this distinction with the actions required for the task within a reasonable time frame.

Through this example, Dennett illustrates the inability of artificial agents to solve the EAFP, since task performance requires both determination of relevance and rapid generation of an appropriate sequence of actions. In summary, this analysis indicates that a solution to the EAFP must involve simulation of certain human cognition capacities, such as real-time assessment of the scope of action side effects (promptness requirement) and effective discrimination between relevant and irrelevant implications (adequacy requirement). Section 4 examines such requirements and the capacity of the salience function of emotions to satisfy them. Below, the two key difficulties of the EAFP are explored.

II.1. *The intra/intercontextual frame problem and the difficulties of the vastness of information/regression*

Among the multiple interpretations proposed for the frame problem, this section distinguishes between the *intracontextual* and the *intercontextual frame problem* (Wheeler 2008). Given a task, the intracontextual version asks how an agent can, within a reasonable time frame, identify relevant actions among *i*) the many available options, as well as *ii*) the possible consequences of those actions. This interpretation highlights the temporal difficulty of exhaustively examining vast amounts of information, given the large number of actions and consequences that require consideration. The resulting computational intractability underlies the *difficulty of the vastness of information.*

This difficulty concerns how a cognitive system with access to an extensive body of information selects what is relevant for task completion within a limited time, that is, promptly. This problem becomes evident when one considers a strategy that requires analyzing every piece of information in the system and evaluating its relevance to the task. Such a strategy is not feasible when the amount of information available to the agent is overwhelmingly large, as in the case of human cognition. Importantly, this limitation does not arise solely from temporal pressure. Tasks such as defusing a bomb or responding to a fire require both successful execution and timely determination of relevance. Efficient adaptation to changing situations therefore becomes essential to avoid paralysis in the face of assigned objectives. At the decision-making level, effective action requires rapid situational adjustment rather than immobility.

The *intercontextual frame problem* relates to how an agent can appropriately identify the context in which it operates. More precisely, it addresses how an agent determines which environmental characteristics count as relevant, among many possibilities, when specifying the situation that requires action. This interpretation of the frame problem introduces another central difficulty of the EAFP, which some AI researchers

regard as a reason for the inability of artificial systems to determine relevance as successfully as humans do.

This limitation, discussed by Dreyfus (2007), refers to what we call the *difficulty of regression*. For this author, epistemological approaches that rely on rules, symbols, and representations for acting in the world require identification of which rules and representations apply in a given situation. This problem becomes more complex because the world undergoes continuous and often unpredictable change. Effective context determination therefore needs maintaining certain assumptions while modifying others.

One might propose programming a system with multiple representations of possible scenarios, together with heuristics that guide action in each case (Minsky 1981). Even under this strategy, and here lies the basis of Dreyfus's pessimism, the problem of selecting the appropriate rule, representation, or heuristic for a specific circumstance persists. Context determination thus remains unresolved. If the EAFP were solved, clarification would follow regarding the mechanism by which one rule or representation determines the applicability of others given particular inputs. Rule selection implies representation selection, which raises a deeper epistemological question: how does an artificial system recognize the appropriateness of one application rather than another?

Lowe (2001) provides an example that illustrates both difficulties. The author invites readers to imagine reading a menu in a restaurant when an alarm suddenly sounds. From the seat, smoke becomes visible through a window located directly in front. In this situation, a decision is required between approaching the fire to intervene or continuing to read the menu. The question then arises whether an artificial agent could act successfully in such circumstances, in terms of both adequacy and promptness. For Lowe, neither the program responsible for reading the menu nor the one designed to address a fire can independently specify the appropriate action, since each operates within a limited domain of competence. This limitation exemplifies the difficulty of regression and undermines guarantees of adequate relevance determination. Moreover, even with a cross-sectional design, the existence of infinitely many possible combinations would hinder rapid action, thereby illustrating the difficulty of the vastness of information and the absence of guarantees of promptness.

This correspondence between the difficulties of the EAFP and the requirements for its resolution plays a central role in the present analysis. This section has clarified the relationship introduced earlier and identified the aspects that require consideration before examining different solutions. In summary, this review shows that the EAFP may remain unsolved when neither difficulty is overcome, may be fully resolved when both difficulties are addressed, or may be partially solved when only one difficulty is overcome. Examination of these difficulties provides the basis for a more rigorous assessment of the resolutive difficulty of the EAFP in the following section.

III- *The resolutive difficulty: The role of emotions and their salience function*

The main objective of this section is to estimate the scope and limits of the salience function of emotions in addressing the EAFP. Existing contributions to this field generally adopt either an ‘optimistic’ or a ‘pessimistic’ stance toward this possibility. This binary classification proves insufficient, as it overlooks the differences in emphasis and internal disagreement within each position. As a result, it fails to capture the range of epistemic degrees proposed for resolving the EAFP. A more informative strategy consists in examining both positions in light of the difficulty of regression and the difficulty of the vastness of information.

Two clarifications are required before proceeding. First, debates on the epistemological relevance of emotions often focus on the various epistemic functions attributed to them, such as motivational force, salience and relevance, access to facts and beliefs, non-propositional contributions to knowledge, and efficiency. This study restricts its analysis to the epistemic function of the salience of emotions, since this function frequently addresses the EAFP and has generated productive debate (Wild 2008; Dohrn 2008; De Sousa 1980).

The second clarification concerns the concept of *emotion*. Given its central role in this work, lack of conceptual clarification could invite objection. Precise definition remains challenging due to the diversity of meanings and functions attributed to emotions. For the purposes of this study, a descriptive approach proves more appropriate than a strict definition. Among the various interpretations compatible with the salience function, emotions arise as responses to stimuli that trigger affective states. Under this understanding, emotions comprise organismic changes produced by internal or external stimuli that possess relevance for the agent. Although this analysis does not specify the extent or type of changes required for emotional occurrence, it highlights two components of the emotional process: motivational and cognitive factors involved in stimulus evaluation (Frijda 1986; 1988).

Emphasis on stimulus significance has led several authors to interpret emotions as *relevance detectors*. Under this interpretation, emotions assist agents in focusing, within a reasonable time frame and with limited cognitive resources, on situational aspects that matter for goal achievement, while avoiding irrelevant considerations. In line with the present interpretation of the EAFP, emotions would therefore contribute to determining relevance both promptly and adequately, and thus successfully. Further specification of the most appropriate conception of emotion would enhance this analysis and remains part of the broader research agenda. The following two sections explore the epistemic success of emotions in relation to each of the main difficulties of the EAFP.

III.1. *The epistemic success of emotions in relation to the difficulty of the vastness of information*

De Sousa ranks among the most optimistic authors regarding the capacity of emotions to resolve the EAFP. According to his view, and within the standard model of rational choice, this aspect arises when an agent confronts a combinatorial explosion of possible choices. In such circumstances, real agents with limited cognitive resources must restrict the range of actions and their potential consequences in order to decide and act successfully. Standard methods of rational choice cannot accomplish this reduction, since they reintroduce the combinatorial problem by requiring evaluation of each consequence of every possible action.

The defining question of the EAFP thus concerns how an artificial agent can navigate this virtually infinite space of possibilities and determine relevance within a reasonable time frame. De Sousa (1980, p. 137) argues that, in such contexts, emotions function as sources of salience, since “they are determined patterns of salience among objects of attention, lines of inquiry, and inferential strategies”. Through this function, emotions delimit relevant actions and consequences, thereby enabling rapid action. The function of the salience of emotions therefore reduces computational demands by addressing the difficulty of the vastness of information. In light of the correspondences established earlier, this form of epistemic success aligns with satisfaction of the promptness requirement.

In line with de Sousa's approach, Damasio (1994) proposed *the somatic marker hypothesis*, which holds that emotions influence the promptness with which relevance is determined, particularly in decision-making tasks. On this basis, authors such as Megill and Cogburn (2005) argue that this theory may prove useful for addressing the difficulty of the vastness of information. Reasoning and decision-making typically presuppose extensive knowledge of the situation, the available courses of action, and their immediate and future consequences. Although this knowledge does not encompass all possible information, it remains sufficiently extensive to risk immobilizing the agent. Effective action therefore requires strategies that enable rapid selection of relevant information.

According to Damasio's hypothesis, an important process occurs prior to explicit reasoning. When an option becomes associated with a negative outcome, the agent experiences an unpleasant bodily feeling. Damasio terms this bodily response a *somatic state*, and, because it marks a mental representation, he describes it as a *somatic marker*. These markers draw attention to unfavorable outcomes and function as automatic alarm signals that warn against dangerous options. When a negative feeling accompanies a potential outcome, the resulting signal often leads to immediate rejection of that course of action, thereby reducing the set of alternatives under consideration. Conversely, positive somatic markers guide the agent toward favorable options.

In decision-making contexts, agents initially consider available options and their consequences, a process that already narrows the field of possibilities but may still leave a large amount of information to evaluate. The somatic marker hypothesis proposes that emotional mechanisms further reduce this information burden. By automatically registering the positive or negative value of possible consequences, somatic markers rapidly filter options without requiring exhaustive evaluation. These mechanisms stabilize motivational tendencies by promoting certain alternatives and discouraging others. Through this filtering process, emotional salience highlights dangerous options and swiftly excludes them, thereby limiting the number of alternatives and consequences that require detailed consideration. As a result, subsequent reasoning operates over a reduced informational space and becomes less demanding. Once again, the epistemic success attributed to emotions corresponds to satisfaction of the promptness requirement.

In summary, this section has examined two optimistic approaches that attribute epistemic success to emotions regarding the difficulty of the vastness of information, albeit with different emphases. Highlighting these differences allows for a more nuanced assessment of the variability within optimistic positions. The following section explores the epistemic success of emotions in relation to the difficulty of regression.

III.2. *The epistemic success of emotions in relation to the difficulty of regression*

Among those who defend, more explicitly, the epistemic success of emotions in the face of the difficulty of regression is Elgin (1996), for whom emotions share several cognitive functions with beliefs and perceptions. Like beliefs, emotions direct our attention toward specific facets of situations. For this author, certain aspects of a situation automatically demand our attention, eclipsing others that are nevertheless epistemically accessible. If a factor is salient, or at least appears to be so, it is presumably significant or relevant.

It is a well-established fact that knowing, understanding, perceiving, or discerning something requires overlooking a great deal of information. However, the question that Elgin considers fundamental is which information must be ignored in order to determine what is relevant and, more importantly, how we know that this selection is made adequately. This latter question directly addresses the difficulty of regression and, correspondingly, the adequacy requirement.

Elgin maintains that salience, or the kind of encapsulation it performs, prevents agents from engaging in a second-order questioning about the rules employed in their decision procedures. From this perspective, determining which rule to apply does not entail asking which further rule should guide that choice. On the contrary, Elgin argues that the salience generated by emotions effectively dissolves the difficulty of regression. Since our capacity to reflect on the role of emotions is limited, agents act according to their emotions without engaging in second-order reflection about how they know that the rule or representation employed is adequate for determining relevance.

This same strategy is also adopted by Hookway (2003), for whom emotions play a much broader role than they do for De Sousa. Hookway not only regards emotions as immediate and indispensable for action and thought, but also assigns them a crucial role in human rationality. According this author, the EAFP questions how we can account, by means of explicit rules, for the way in which humans agents infer inductive generalizations from singular sensory stimuli. Strategies for determining relevance in a Quinean sense (Quine 1960) presuppose the capacity to conduct an exhaustive and accurate search across a vast body of information. Since no one knows how to model a system capable of meeting these conditions, we are unable to explain the explicit rules governing this process without falling into regression; consequently, the EAFP remains unsolved.

This interpretation, which Hookway shares with Dreyfus, emphasizes the need to provide explicit rules for determining relevance without thereby generating infinite regression. Both authors thus address what is commonly referred to as the difficulty of regression, although Hookway places particular emphasis on its inductive dimension. For this reason, he connects this difficulty with the well-known *grue* paradox (Goodman 1983, p. 74). This paradox captures the problem of induction: having observed a sufficiently large number of green emeralds, we tend to infer that all emeralds are green. However, we could instead define the predicate *grue*: an object is *grue* if and only if it has been observed to be green so far or has not yet been observed to be blue. Goodman (1983) asked why we infer from our observations that all emeralds are green, but not that all emeralds are *grue*.

In response, Hookway argued that, although we cannot articulate explicit rules explaining why the former inference is more compelling than the latter, we nonetheless have reasons to experience it as such. In other words, he maintained that neither the how nor the why of this preference is something we can fully grasp through reflective understanding. One of the reasons why certain generalizations, such as ‘emeralds are green’, strike us as more convincing than others, according to Hookway, lies in an affective or emotional response that resists detailed explanation. We immediately feel comfortable with the claim that emeralds are green, whereas the opposite reaction arises in the case of ‘emeralds are grue’. Hookway thus emphasized that the fundamental role played by inductive generalizations in our epistemic practices cannot be entirely accounted for at a reflective level. Far from indicating a limitation or failure of rationality, this situation shows that it is rational, appropriate, and epistemically responsible to rely on inductive generalizations grounded in such affective responses.

Briefly, the authors discussed above adopt an overall ‘optimistic’ stance when defending the role of emotions with respect to the difficulty of regression and, correspondingly, in securing epistemic success in relation to the adequacy requirement. The following section turns to the pessimistic accounts and examines the scope of epistemic failure.

III.3. *Epistemic failure in relation to the difficulty of regression*

In contrast to epistemological positions that defend a certain epistemic success in appealing to emotions, other accounts at least cast doubt on such success. Let us begin by examining Dohrn's (2008) critique of the salience function of emotions as a potential solution to the EAFP. According to Dohrn, endorsing this strategy ultimately amounts to appealing to emotions in order to justify why Dennett's robot fails to complete its task, namely defusing a bomb. From this perspective, he emphasizes that de Sousa suggests that, had the robot been guided by emotions rather than reason, it would have been able to determine relevance adequately. On this view, emotions would appear to be more reliable in determining relevance, insofar as they would allegedly lead only to correct conclusions.

Dohrn challenges this assumption by asking why determining relevance through emotions should yield better results than doing so through a random process. He argues that emotions are invoked merely to 'fill the gap' left by reason: since Bayesian decision theory, understood as a rational framework, cannot account for what an agent ought to do in situations of indifference—where several mutually exclusive options have equal expected utility—emotions are introduced to resolve the impasse. However, Dohrn asks why this gap could not just as well be filled by appealing to some other mechanism, such as a process governed purely by chance. More importantly, he questions what criteria could justify the claim that emotions would perform this task more successfully.

If the sole function of emotions were to 'fill in' what reason fails to determine with respect to relevance, then according to Dohrn, virtually any alternative explanation could be admitted. The core of his criticism is that, if emotions merely compensate for the shortcomings of rational procedures, there is no principled way to evaluate their epistemic success. Returning to central issue, if emotions are invoked only because reason is unable to resolve the problem, then the claim that they make a distinctive contribution to satisfying the adequacy requirement—and thus to overcoming the difficulty of regression—lacks sufficient grounding. In other words, emotions would be required precisely because reason fails, not because they provide an independently justified epistemic advantage. From this standpoint, the epistemic success of emotions cannot be legitimately invoked in addressing the difficulty of regression, since their role is limited to that of a stopgap or filler.

Based on Wild's (2008) critiques, according to which the EAFP can be interpreted as a decision problem, there are procedures other than emotional salience that also reduce the epistemic search space by selecting among alternatives within a reasonable time. These procedures include tossing a coin; consulting friends, colleagues, or experts; prioritizing interests and objectives; choosing means that are good enough relative to those objectives and interests; and employing quick and frugal heuristics such as 'never change a winning team', 'better safe than sorry', or 'take the first', among others. The frame problem is sometimes understood as questioning whether it is reasonable to examine 'all' possible alternatives, as well as all the consequences associated with each of them. Clearly, such an exhaustive examination would be highly irrational.

Nevertheless, emotions are not decisive by themselves. Rather, they constitute 'part of the material' that comes into consideration when agents face the problem of how to delimit the epistemic space or choose between various alternatives, without being exclusive to the decision-making process. In some cases, emotions and other affective states merely interfere with other decision-making strategies or otherwise reasonable judgments, and may even take direct control of the process. In this sense, Wild acknowledged that emotions contribute to addressing the difficulty of the vastness of information, but he questioned whether they do so either exclusively or correctly. Although selecting certain information may be appropriate, there may be several other rea-

sons that prevent us from choosing it. In such cases, agents decide and act without reflection. Here, emotions take control only momentarily, insofar as they deactivate other dispositions or judgments that are temporarily weakened. Consequently, there is no guarantee that emotions can determine relevance adequately.

Returning to our central concern, from this perspective the salience function of emotions as a means of resolving the difficulty of regression is at least weakened. This is because it is doubted not only that appealing to emotions guarantees an adequate determination of relevance, but also that such a task is exclusive to emotions.

Goldie (2008) shares a similar perspective on the role of emotions in decision-making, although he is more sympathetic to their capacity to solve certain outstanding problems. For Goldie, emotions can distract us from epistemic understanding, since they may still 'skew the epistemic landscape'. He warned of an overly optimistic trend in contemporary epistemological theories on emotions, a trend reinforced by their characterization as 'quick and frugal heuristic' within the framework of bounded rationality. According to Goldie, this optimism overlooks the possibility that emotions may systematically mislead us.

Let us consider one final criticism, which concerns the *globality* of emotions. On the basis of the foregoing discussion, emotions appear to motivate and guide epistemic activities: they are intrinsically evaluative, possess a distinctive phenomenal character, regulate inquiry, and assess both activities and articulated norms, among other functions. In short, affective states seem to perform an excessive amount of epistemological work. Wild compared this situation to a film in which a single actor plays several main characters simultaneously. The problem with such an expansive attribution of roles to emotions is that they come to occupy too many fundamental functions, which, at least for Wild, and in line with Dohrn's criticism of de Sousa, reduces their role to a mere 'filling function' for processes that cannot be explained by reason alone.

According to Wild, the proliferation of theories and the presence of conflicting positions regarding the epistemic role of emotions hinder the development of a unified theory of emotion, thereby justifying a certain degree of skepticism about their theoretical usefulness. Moreover, even if we were to delimit the specific cognitive tasks attributed to emotions, the question of whether *emotion* constitutes a uniform and explanatory useful theoretical concept would still remain open. Addressing these issues exceeds the scope of the present work, although they remain an important task within the interdisciplinary domain that we have delineated.

Briefly, the authors discussed in this section predominantly adopted a 'pessimistic' position regarding the scope of emotions in relation to the difficulty of regression, though not with respect to the difficulty of the vastness of information. The following section presents the main findings of the present investigation, assessed in terms of the degrees of epistemic success.

IV. *Degrees of epistemic success*

From an optimistic position, we have seen that for both de Sousa and Damasio, emotions contribute to classifying certain options as relevant and to excluding others from such consideration. In other words, emotions determine salience by circumscribing practical and cognitive options within a vast informational space. Thus, both authors attribute epistemic success to emotions with respect to the difficulty of the vastness of information. Along similar lines, Elgin and Hookway argued that the salience function of emotions, in addition to participating in this task, ensures that it is carried out successfully. Therefore, emotions also contribute to resolving the difficulty of regression.

As previously noted, the differences within this broadly optimistic position concern the degree of emphasis placed on the epistemic success achieved in relation to each of the two difficulties. By contrast, from a more pessimistic perspective, we examined several criticisms directed at the so-called filling function of emotions. These objections primarily question why emotions should be expected to determine relevance *reliably* and *appropriately*. They do not challenge the role of the emotional salience in reducing informational complexity, but rather this reduction is achieved in a normatively adequate manner. In other words, they cast doubt on the scope of emotions in resolving the difficulty of regression, while leaving their role in addressing the difficulty of the vastness of information largely intact.

Given the requirements identified earlier, the debate can be understood as centering on the issue of *guarantee*. As the analysis suggests, the epistemic success of emotions appears to be guaranteed in relation to the requirement of promptness, insofar as emotions effectively reduce the amount of information to be processed. However, such success is not guaranteed with respect to the requirement of adequacy, since it remains doubtful that emotions reliably discard only what is epistemically irrelevant. Consequently, greater epistemic success is attributed to emotions in resolving the difficulty of the vastness of information than in addressing the difficulty of regression. While there is broad agreement that the salience function of emotions contributes to narrowing the informational field, it is not established that this narrowing is carried out *certainly* and *appropriately*.

In the literature, the relative importance attributed to the scope of emotions with respect to each of the difficulties of the EAFP, considered separately, has often been the source of serious misunderstandings. This research has addressed that gap by examining epistemic aspects without committing to either extremely optimistic or extremely pessimistic positions. The features highlighted here focus specifically on the salience function of emotions; however, other lines of inquiry have concentrated on different epistemological issues, such as the types of emotions involved (Jones 2007). Concerning our interface area, identifying and distinguishing these aspects supports the relevance of reapproaching the EAFP through lateral epistemological considerations, such as *a)* conceptualizations, *b*) classifications, and, as undertaken here, *c*) distinct epistemological functions of emotions. Clarifying these issues from the perspective of the Epistemology of Emotions would also contribute to assessing the scope of the salience function of emotions in relation to other central problems in AI.

Before proceeding, it is useful to briefly return to Dennett's robots, now in light of the foregoing analysis and under the assumption that emotions could be attributed to an artificial agent. At its core, the EAFP asks whether an artificial agent endowed with emotions could determine relevance as successfully as humans do. According to our analysis, this question gives rise to two further considerations. The first concerns whether, through the salience function of emotions, an artificial agent could determine relevance with a degree of promptness comparable to that of humans. Addressing this issue requires estimating the degree of epistemic success of emotions with respect to the requirement of promptness in resolving the difficulty of the vastness of information. In this sense, task success would depend on the extent to which emotions contribute to acting quickly enough to avoid catastrophic outcomes, such as the explosion of the bomb.

The second consideration concerns whether an artificial agent could determine relevance with a degree of adequacy comparable to that typically exhibited by humans. Answering this question requires evaluating the degree of epistemic success of emotions with respect to the requirement of adequacy in resolving the difficulty of regression. In this case, successful task completion would depend on whether emotions contribute to

discriminating reliably between information that is genuinely relevant and information that should be excluded as irrelevant.

Establishing such discrepancies, grounded in specific epistemological clarification, underscores the importance and scope of the relationship—introduced at the outset of this paper—between the Epistemology of Emotions and AI. When confronted with this challenge, epistemologists and researchers in AI converge, whether explicitly or implicitly, on a shared objective (Davidson et al. 2009): to explain the inherently rational nature of emotions—namely, their capacity to support rational inference, evaluation, and estimation of information—with the aim of optimizing these capacities or developing artificial systems that emulate them. Nevertheless, and as this work highlights, the question remains: in what epistemological terms is this challenge framed? When comparing artificial and human emotions, which contexts are presupposed, and which conceptions, types, and functions of emotions are taken into account? In this regard, we have argued for the need to re-examine this issue from a specific epistemological perspective.

Pinker (2005, p. 10) noted that, according to Fodor (2000), the frame problem ultimately reflects "an unbridgeable gulf between the feats of human abduction and the powers of computational systems". In response, we have questioned whether this claim should be interpreted as a strong mathematical thesis—namely, that a certain function cannot, in principle, be computed by a Turing machine—or rather as the identification of a function that computational systems cannot perform with the same degree of speed and adequacy as humans.

This present study is limited in scope, as it addresses only one among several central problems in AI research. In this sense, our analysis is necessarily partial and leaves numerous questions open. Nevertheless, our findings may contribute to a more precise assessment of the prospects for developing sociable robots, machines, and interfaces that establish emotional relationships with users through designs optimized in terms of adequacy and/or promptness (Breazeal 2002). They may also inform debates in other domains where epistemic failures have been anticipated, such as concerns surrounding the algorithmic erosion of democratic processes. In this context, Innerarity (2023) has articulated a form of epistemological pessimism regarding politics, understood as a domain that does not operate through linear or deductive reasoning but involves managing deeply ambiguous situations and making rapid yet appropriate decisions under conditions of uncertainty and contingency—conditions that contrast sharply with the demands of algorithmic logic for clarity, objectivity, and precision. In this broader landscape, the Epistemology of Emotions may contribute to the development of theoretical models applicable to virtual environments that promote epistemically successful action (Mun 2019).

V. *Conclusion*

This work has addressed, from a multidisciplinary perspective and within the context of the frame problem, the linkage between cognition, emotion, and decision-making. More specifically, our findings support the significance of the salience function of emotions in overcoming the main difficulties of the EAFP, which broadly concerns how successfully an artificial agent can determine relevance. The analysis suggests that emotions exhibit greater epistemic success in resolving the difficulty of the vastness of information (promptness requirement) than in resolving the difficulty of regression (adequacy requirement).

In summary, this study has raised key questions regarding: *i*) the nature of the difficulty of the vastness of information and the difficulty of regression from a cognitive perspective; *ii)* the aptitude of the salience function of emotions to address each of these difficulties independently; and *iii)* the different degrees of epistemic success achieved in each case with respect to decision-making. Clarifying these issues contributes to a more rigorous assessment of the implications of adopting an epistemological perspective on emotions for their potential practical application in the advancement of AI systems grounded in the interaction between *cognition, emotion*, and *decision-making.*

Further research is required to determine: *a*) how other aspects of the frame problem may be affected by emotional salience; *b)* which additional epistemological functions could contribute to its resolution; and *c*) the range of contexts to which the present proposal may be extended. Moreover, by restricting our analysis to a single approach to resolve the EAFP rather than addressing the problem in its entirety, this work should be understood as part of a broader research program, which we intend to continue developing in future studies.

VI. *Bibliographic references*

ASSUNCAO, G., PATRAO, B., CASTELO-BRANCO, M., & MENEZES, P. (2022), "An Overview of Emotion in Artificial Intelligence", *IEEE Transactions on Artificial Intelligence*, pp. 867-886.

BREAZEAL, C. (2002), *Designing Sociable Robots*, Cambridge: MIT Press.

BROWN, F. (1987), The frame problem in artificial intelligence, *Proceedings of the 1987 Workshop*, Los Altos: Morgan Kaufmann.

BRUN, G., & KUENZLE, D. (2016), "A new role for emotions in epistemology", in G. Brun, U. Doguoglu, & D. Kuenzle (eds.), *Epistemology and emotions*, Alderchot: Ashgate, pp. 1-31.

CALVO R., D'MELLO S., GRATCH J., & KAPPAS A. (2015), *The Oxford handbook of affective computing*, New York: Oxford University Press.

CANDIOTTO, L. (2019), "From Philosophy of Emotion to Epistemology: Some Questions About the Epistemic Relevance of Emotions", in L. Candiotto (ed.), *The Value of Emotions for Knowledge*, London: Palgrave Macmillan, pp. 3-24.

CROCKETT, L. (1994), *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence*, New Jersey: Ablex Publishing Corporation.

DAMASIO, A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Putnam.

DAVIDSON, R., SHERER, K., & GOLDSMITH, H. (2009). *Handbook of affective sciences*, New York: Oxford University Press.

DE SOUSA, R. (1980), "The rationality of emotion", in R. Amelie (ed.), *Explaining emotions*, Berkeley: University of California Press, pp. 127-151.

DENNETT, D. (1984), "Cognitive wheels: The Frame Problem of AI", in C. Hookway (ed.), *Minds, machines & evolution*, Cambridge: Cambridge University Press, pp. 129-152.

DEONNA, J., & TERONI, F. (2012), *The Emotions: A Philosophical Introduction*, London and New York: Routledge.

DOHRN, D. (2008), "Epistemic immediacy and reflection", in G. Brun, U. Doguoglu and D. Kuenzle (eds), *Epistemology and emotions*, Alderchot: Ashgate, *pp.* 105-124.

DREYFUS, H. (2007), "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian", *Artificial Intelligence*, 20, 2, pp. 1137-1160.

ELGIN, C. (1996), *Considered judgment*, Princeton: Princeton University Press.

EVANS, D. (2004), "Can robots have emotions?", *Psychology Review*, 11, pp. 2-5.

FODOR, J. (1983), *The modularity of mind*, Cambridge: MIT Press.
FODOR, J. (2000), *The mind doesn't work that way: The scope and limits of computational psychology*, Cambridge: MIT Press.
FRIJDA, N. (1986), *The emotions*, Cambridge: Cambridge University Press.
FRIJDA, N. (1988), "The laws of emotion", *American Psychologist*, 43, pp. 349-358.
GLYMOUR, C. (1987), "Android epistemology and the frame problem: Comments on Dennett's cognitive wheels", in Z. Pylyshyn (ed.), *The robot's dilemma: The frame problem in artificial intelligence*, Norwood: Ablex Publishing Corporation, pp. 65-77.
GOLDIE, P. (2008), "Misleading emotion", in G. Brun, U. Doguoglu and D. Kuenzle, (eds.), *Epistemology and emotions*, Alderchot: Ashgate, pp. 149-167.
GOODMAN, N. (1983), *Fact, Fiction, and Forecast*, Cambridge: Harvard University Press.
HJØRLAND, B. (2010), "The Foundation of the Concept of Relevance", *Journal of the American Society for Information Science and Technology*, 6, 2, pp. 217-237.
HOOKWAY, C. (2003), "Affective states and epistemic immediacy", *Metaphilosophy*, 34, 1/2, pp. 78-96.
INNERARITY, D. (2023), "The epistemic impossibility of an artificial intelligence take-over of democracy", *AI & Society*, 10, pp. 1-5.
JONES, K. (2007), "Quick and Smart? Modularity and the Pro-Emotion Consensus", *Canadian Journal of Philosophy*, 36, 5, pp. 3-27.
KAPLAN, A., & HAENLEIN, M. (2019), "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business horizons*, 62, 1, pp. 15-25.
LOWE, E. (2001), *Philosophy of the mind*, Barcelona: University Idea.
MAJEED, R. (2020), "Does modularity undermine the pro-emotion consensus?", *Mind & Language*, 35, 3, pp. 277-292.
MCCARTHY, J., & HAYES, P. (1969), "Some philosophical problems from the standpoint of AI", *Machine Intelligence*, 4, pp. 463-502.
MEGILL, J., & COGBURN, J. (2005), "Easy's getting harder all the time: Human emotions and the Frame Problem", *Ratio*, XVIII, 3, pp. 306-316.
MINSKY, M.(1968), *Semantic information processing*, Cambridge: MIT Press.
MINSKY, M. (1981), "A Frame problem for Representing Knowledge", in J. Haugeland (ed.), *Mind Design*, Cambridge, The MIT Press, pp. 95-128.
MUN, C. (2019), "How Emotions Know: Naturalizing Epistemology via Emotions", in L. Candiotto (ed.), *The Value of Emotions for Knowledge*, Cham: Palgrave Macmilla, pp. 27-50.
PICARD, R. (2000), *Affective Computing*, USA: MIT Press.
PINKER, S. (2005), "So how does the mind work?", *Mind & language*, 20, 1, pp. 1-24.
POLITOU, E., ALEPIS, E., & PATSAKIS, C. (2017), "A survey on mobile affective computing", *Computer Science Review*, 25, pp.79-100.
QUINE, W. (1960), *Word and Object*, Cambridge: MIT Press.
RABOSSI, E. (2002), "Philosophy of mind and philosophy of psychology: the agenda, the practice, the domain", *Azafea.*, 4, pp. 21-43.
WHEELER, M. (2008), "Cognition in context: phenomenology, situated robotics and the frame problem", *International journal of philosophical studies*, 16, 3, pp. 323-349.
WILD, M. (2008), "Critical reflection on affective epistemology", in G. Brun, U. Doguoglu and D. Kuenzle (eds.), *Epistemology and emotions*, Alderchot: Ashgate, pp. 125-149.

Maria Ines Silenzi es investigadora en el Instituto de Investigaciones Económicas y Sociales del Sur (IIESS/CONICET) y profesora en el Departamento de Humanidades, Universidad Nacional del Sur, Argentina.

Líneas de investigación:
Su línea de investigación se orienta hacia la filosofía de la mente en general y en la epistemología de las emociones en particular, en intersección interdisciplinaria con algunos problemas claves dentro de la Inteligencia Artificial.

Publicaciones recientes:
- (2025). "The frame problem: Relevance, emotion, and degrees of epistemic success." *Rivista internazionale di Filosofia e Psicologia*, 16(1), 49-65.
- (2025) "Sobre toma de decisiones y procesos de relevancia: la función epistemológica de saliencia/relevancia de las emociones en agentes económicos", *Revista latinoamericana de Filosofía* 51 (1), 77-105.

Email: misilenzi@uns.edu.ar