DOI: 10.20911/21769389v51n161p527/2024

IS THERE A PLACE FOR EMOTIONS IN SOLUTIONS TO THE FRAME PROBLEM? *

Há Espaço para Emoções em Soluções do "Frame Problem"?

Carlos Barth **

Abstract: **The frame problem, a long-standing issue in Artificial Intelligence (AI), revolves around determining the relevance of information in an ever-changing array of contexts, posing a formidable challenge in modeling human reasoning. The purpose of this paper is to explore the hypothesis that emotions are able to solve, or at least enable a substantial step towards a solution. I argue that, while emotions are integral to cognitive processes, they do not offer a solution to the frame problem, nor can they play a significant role in a solution. In the course of the argument, I'll show why some previous claims that emotions could be helpful towards a solution rely on a problematic analysis of what the frame problem is. Ultimately, the conclusion is that no context-dependent feature can play a non-trivial role in solutions to the frame problem, and that is the case of emotions.**

Keywords: **Frame problem. Emotions. Artificial Intelligence. Cognition. Relevance.**

Resumo: **O "frame problem" é o desafio de determinar a relevância de uma informação em um conjunto fluido e indefinidamente multiplicável de contextos, representando um desafio formidável para a modelagem do raciocínio humano como almeja a Inteligência Artificial (IA). O propósito deste artigo é explorar a hipótese de que as emoções são capazes de resolver, ou pelo menos possibilitar um avanço substancial em direção a uma solução. Argumento que, embora as**

* Artigo recebido em 23.10.2024 e aprovado para publicação em 07.11.2024.
** PhD in Philosophy from the Federal University of Minas Gerais (UFMG). Postdoctoral researcher in the Department of Philosophy at the Jesuit Faculty of Philosophy and Theology (FAJE). E-mail: carlos@cbarth.me

emoções participem dos processos cognitivos, a resposta é negativa. No decorrer do argumento, mostrarei por que algumas afirmações anteriores de que as emoções poderiam ser úteis para uma solução dependem de uma análise problemática do que é o "frame problem". Em última análise, a conclusão é que nenhum recurso dependente de contexto pode desempenhar um papel não trivial em soluções para o "frame problem", e esse é o caso das emoções.

Palavras-chave: **Frame problem. Emoções. Inteligência Artificial. Cognição. Relevância.**

Introduction

Commonsense usually pictures feelings and reason as antagonists. The ability to prevent one from encroaching on another's space is generally regarded a virtue. But what exactly could amount to a "space of reasons" and a "space of feelings", assuming there are such spaces? And what could their relation be? These are hard questions, specially considering the richness and variety of the affectivity phenomenon. For instance, nausea, fear and being in love are different kinds of feeling that render distinct effects on intellectual activity.

This issue was inherited by the cognitive sciences. How should we account for the possible effects of feelings within cognitive processing? For decades, this question was given very little space. A typical assumption is that one can safely ignore full-fledged accounts of feelings, for cognitive science only needs to worry about their effects *qua* cognitive assets. Of course, what exactly comprises a "cognitive asset" revolves around the set of tools yielded by one's favorite explanatory framework. For instance, in classic symbolic approaches, feelings would matter only inasmuch as they produce symbols that play a role in the system's reasoning. Things like "fearful spider here now" or "right toe hurts". Other aspects of such felt qualities would be simply put aside as somebody else's business. Such segregative stance is quite appealing, for it avoids the odd conclusion that our feelings have no influence on our thinking while keeping affectivity-related issues out of the way.

However, as the years passed, the interest for the affectivity dimension has grown, and with it the need for more comprehensive and detailed accounts. For instance, Minsky[1] suggests that we can account for feelings through the functional analysis of our cognitive machinery. In this view, being angry could be analyzed into activating resources that allows one

[1] MINSKY, M. The Emotion Machine: *Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006.

to act with more speed and strength while mitigating resources related to one's cautiousness. However, more recent trends in cognitive science reject this approach. As an example, Mesquita and Boiger[2] propose a sociodynamic model in which emotions emerge in the context of social interactions. They are, in this sense, interpersonal systems. In that case, being angry with one's boss is a distinct feeling from being angry with one's child.

With this shift in perspective, researchers began to reconsider long-standing issues and hypothesized that an understanding of feelings might be of help. This is the case of the *frame problem* (FP), an enduring hurdle in modeling human reasoning. It presents a significant challenge in handling tasks that come naturally to humans, such as realizing what's relevant in an open--ended set of contexts. Although the issue was extensively discussed for some decades, there was no significant progress.[3] The lack of progress led to the impression that a fundamental piece was missing. The renewed interest in feelings made them an obvious candidate. Hence, in the course of the debates about the frame problem, researchers like Sousa begin to believe that certain types of feelings, particularly emotions, might play a crucial role in addressing it.[4]

Of course, if feelings are to bring something original to the table, they must offer more than just cognitive assets like those assumed by the segregative stance. If we reduce the role of feelings to that of mere information to be processed, that won't buy us anything new. Furthermore, feelings can't be seen as mere inspiration for heuristics or algorithms. Even if such inspiration led to exciting results, that would be an achievement of the formulated algorithm, not the feeling that inspired it.

To develop this idea in full, we would need a comprehensive understanding of what feelings are, but such an account is not currently available. Therefore, we will focus on the cognitive roles that feelings must play in order to be considered a solution (or a partial solution) to the frame problem.

I can already state that the conclusion of this work is rather pessimistic, though. I'll argue that emotions cannot offer neither a complete solution to the frame problem, nor a substantial step towards one. The details matter, so here's how I'm going to unpack this claim. First, I want to narrow down the idea of feelings to the narrower concept of emotions. As I understand the term, "feeling" encompasses everything from simple sensations (taste

[2] MESQUITA, B.; BOIGER, M. Emotions in Context: A Sociodynamic Model of Emotions. *Emotion Review*, v. 6, n. 4 (2014) p. 298-302.
[3] See FODOR, J. A. Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In: PYLYSHYN, Zenon W. (Ed.). *The robot's dillemma: The frame problem in artificial intelligence.* Norwood, NJ: Ablex, 1987. p. 139-149. See also FORD, K. M.; PYLYSHYN, Z. W. (EDS.). *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence.* Norwood, NJ: Ablex, 1996.
[4] SOUSA, R. *The Rationality of Emotion.* Cambridge: MIT Press, 1987.

of salt, warmth, thirst, etc.) to whole existential moods characterizing our stance towards the world (being depressed, joyful, seeing the world as a scary place, etc.). Here I follow Haugeland[5] and reserve "emotion" for things that are neither simple sensations nor existential moods (or simply moods). I mean occurring episodes of feelings like anger, fear, or gratitude. The goal is not to provide a clear-cut distinction, nor a full--fledged taxonomy, though. Only to differentiate emotions from feelings that, despite having a felt quality, have no clear valence, such as taste of salt. The distinction between emotions and moods is equally important. While emotions such as "fear of a spider" bear an effect on the agent's behavior in certain occasions, moods have a somewhat global character and has a large-scale influence on how the agent is poised to handle the world. I have more to say about the possible role of moods, but only after fully developing the core argument regarding emotions.

With that in mind, in section 2 I'll present my reading of the frame problem, making a clear demarcation between two common interpretations and stating the one we'll adopt here (the frame problem as an issue about relevance determination). In section 3, I'll present a framework (somatic markers model, SMM) in which we'll articulate the possible roles of emotions in cognitive processes. After that, in section 4, I'll present the core argument by discussing the works of Ransom[6] and Wheeler[7]. In particular, I'll show that the idea that emotions may help make a significant step towards a solution only seems plausible due to a problematic analysis of the frame problem. Finally, I'll proceed to some closing remarks.

1. What's the frame problem about?

The frame problem first emerged in classical artificial intelligence (AI) as a significant challenge in modeling the trajectory of human reasoning.[8] In early classical AI, the *logicist* approach was prevalent. Its goal was to model the system's line of reasoning with a logical formalism. The particular formalism developed by Hayes and McCarthy's was named *situation calculus*.[9] Roughly, their approach involved modeling the current state of

[5] HAUGELAND, J. *Artificial Intelligence: The Very Idea.* Cambridge, MA: MIT, 1985.
[6] RANSOM, M. Why Emotions Do Not Solve the Frame Problem. Em: MULLER, Vincent (Ed.). *Fundamental Issues of Artificial Intelligence*. Berlin: Springer, 2016, p. 353-365.
[7] WHEELER, M. Cognition in context: phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, v. 16, n. 3 (2008), p. 323-349.
[8] MCCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, v. 4 (1969), p. 463-502.
[9] For a formal introduction to situation calculus, see SHANAHAN, M. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* Cambridge, MA: MIT, 1997.

affairs of the agent's world as a set of axioms. In this set we would find sentences specifying objects, properties and relations among them. There could be axioms expressing "there's a cup", "there's a table", "the cup is empty", "the cup is on the table at position x,y", and so on. Whenever some action is executed, or some event happens, the situation changes from s1 to, say, s2. Thus, if one puts coffee on the cup, the description of the agent's world must be updated. The sentence "there's a cup" was true in s1, and it is still true in s2, but "the cup is empty" is not true anymore in s2. Now, if we try to figure out these updates ourselves, it seems like a rather tedious yet manageable endeavor. The goal, however, is an automated system that is able to do such updates on its own. And that's where the real challenge begins.

The core of the problem is: given an action or event, how can the system decides what changes and what stays the same? It all depends on the event we're talking about. In situation calculus, an event (or action) is modeled as a set of rules specifying what to update. Therefore, an event like "coffee was put on the cup" comprises rules such as "the cup becomes non-empty", "the bottle from which the coffee came from becomes less full", and so on. By following these rules, the system can figure out the updates by itself. If someone asks it something like "if I put coffee in the cup, will the cup's color stay the same?". The system can verify its rule set and, provided that there's no rule saying anything about cups changing color when coffee is put in them, it replies with "yes".

But that's clearly not enough. Assume that the cup in question is transparent and that we can see through it. Were we to ask the system "If someone puts coffee in the cup, will I still be able to see through it?" It's answer would be yes, even though that's evidently false. What the system needs is another rule stating that putting coffee in a cup won't change the cup's color *unless* the cup is transparent. Other possibilities must be accommodated as well, so we'd probably need additional clauses like: *unless* it is made of thermochromic material, or *unless* is has some head-sensitive coating, or *unless* its surface triggers a chemical reaction, and so on. There's more, though. What goes for color also goes for properties such as shape, weight and temperature. Putting coffee in a cup won't change the cup's shape, *unless* the cup is made of thin plastic (just like disposable ones) and the coffee's temperature is too high. Putting coffee in a cup will change the cup's temperature, *unless* the coffee is already at room temperature, or *unless* the cup is made up of low thermal conductivity material, or *unless* the cup was preheated, and so on.

As we add more objects, properties and relations to the system's cognitive world, i.e. as we render it more realistic, the number of necessary clauses increases very quickly, and if the number is too large, the system becomes unmanageable.

The upshot is that whenever we say that putting coffee in a cup won't change the cup's color, we're not saying that's always the case. Rather, we're stating that's *ordinary*. In ordinary conditions, whatever that amounts to, putting coffee in a cup won't change the cup's color. Somehow we're able to stick to what's ordinary without losing sight of what's extraordinary yet possible.[10] But how can we model ordinariness, other than as a prohibitively large set of unless-like clauses? How can we draw a principled frame distinguishing what's ordinary and what's not? That is what McCarthy and Hayes dubbed the frame problem.[11]

Once philosophers became aware of the frame problem, a kind of war of interpretations broke out. Rather than ordinariness, Dennett[12], Haugeland[13] and Fodor[14] took the issue to be about *relevance*. Currently, most of the literature dedicated to the frame problem (if not all), understand it as the difficulty in realizing what's relevant in an open-ended set of contexts. It might seem that figuring out what's relevant would be enough to handle what's ordinary as well. After all, if one sticks to what's relevant, everything needed seems to be already in place. When trying to figure out the consequences of an action or event, the system will consider every circumstantially relevant possibility, ordinary or not.

Tempting as it is, I think this reasoning must be rejected. While the need for ordinariness is about modeling change, the need for relevance sensitivity is about exploiting cognitive assets. Perhaps an example can highlight the distinction. Say you are an avid reader and often make summaries of your readings. Then someone asks you a question about chess. Luckily, you have a well organized set of notes from your readings. As you hear the question, you know that the box with notes about chess will be relevant

---

[10] Notice one should not put too much weight on what "extraordinary" means. There's nothing really exceptional about putting room-temperature coffee on one's cup (by accident or malice). The point is that whenever we handle or talk about situations involving cups of coffee, this is the kind of circumstance we expect to be made explicit. In ordinary situations, no one expects to be served cold coffee without being warned about it.

[11] This is evidently a very simplified tale that sacrifices historical accuracy for the sake of clarity and simplicity. Situation calculus has no support for anything like an "unless" logical operator. Instead, events and actions are mapped by effect axioms and frame axioms. The former describes the effects of an action or event (temperature gets changed), and the latter describes its non-effects (shape remains, location remains, color remains…). But this means that every single event requires a frame axiom for every single property in the system's world. That's how the frame problem was first described. Though the story here is simplified, it is (hopefully) faithful to Hayes' interpretation of the frame problem. See MCCARTHY and HAYES, Some philosophical problems from the standpoint of artificial intelligence.

[12] DENNETT, D. Inteligência artificial como filosofia e como psicologia. Em: *Brainstorms: ensaios filosóficos sobre a mente e a psicologia*. Tradução de Luiz Henrique De Araújo Dutra. São Paulo: Unesp, 1999[1978], p. 163-183.

[13] HAUGELAND, J. An overview of the frame problem. Em: PYLYSHYN, Zenon W. (Ed.). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex, 1987, p. 76-93.

[14] FODOR, Modules, frames, fridgeons, sleeping dogs and the music of the spheres.

in formulating an answer, so you take it and ignore the other boxes. But then comes another question: do you think that apes could learn how to play rudimentary chess? This is trickier, for there's nothing about ape's cognitive capacities in the chess box. You know you've read about this stuff a while ago, and you know the relevant information is *somewhere*. But it is not organized in the way you need it to be in order to quickly answer the current question. You have no box named outline-of-the-cognitive--capacities-required-to-learn-chess-and-an-evaluation-of-these capacities-in--apes. You do have boxes about apes and cognitive capacities in general, but the contents of such boxes have a lot of information that is irrelevant for the current task. Thus, even if you do have the capacity to recognize what's relevant, you still would have to check the contents of those boxes, one by one (this is relevant, this is not, this is relevant *unless* it happens to be the case that….). The upshot is that, although what's ordinary and what's relevant are both context dependent, what's ordinary in a context is different from what's relevant to one's current purposes.[15]

Ordinariness is about making information retrieval manageable. The importance of this issue is quite clear in broadly representational frameworks. However, its impact on non-representational accounts is less obvious. Non-representational accounts typically avoid terms like "information processing," instead suggesting that systems "tune" themselves to their environment through their skill set. Nonetheless, this skill set must still be stored within the system somehow. Thus, even if a preferred framework doesn't focus on how this know-how is encoded within the system, addressing it can still be beneficial.

Relevance, in turn, is about knowing what to retrieve in a given circumstance. The challenge lies in determining what must be retrieved (i.e., what is circumstantially relevant) before actually doing so. Otherwise, intractability becomes a threat, as the number of possible inferences or features to consider in any situation is simply too vast to manage through exhaustive consideration. There must be a point where one stops deliberating about the multiple ramifications of possible outcomes and takes action, but deciding when to stop is a matter of deliberation itself. This is what Fodor famously labeled "Hamlet's problem from the point of view of an engineer": when to stop thinking?[16]

In current literature, the frame problem is often discussed as a problem of relevance, specifically, Fodor's engineer's version of Hamlet's problem. The frame problem as an issue of ordinariness is sometimes considered solved,

---

[15] Another way to state the same point is that the pursuit or ordinariness is a pursuit of the right ontology. A system that stores information in a way that respects what's regarded as ordinary is a system that is able to carve out the world in the right joints.

[16] FODOR, Modules, frames, fridgeons, sleeping dogs and the music of the spheres.

leaving only the relevance problem standing.[17] However, it is crucial to distinguish between these problems to avoid errors such as those made by Churchland.[18] In that work, Churchland conflated the two problems and claimed that switching from language-like representations (e.g., Fodor's language of thought) to structural representations within neural networks could fully avoid the frame problem. This is mistaken. While structural representations may help with the issue of ordinariness, they do nothing to address the issue of relevance.[19]

The resulting picture contradicts the view that solving the relevance issue would simultaneously solve the frame problem as an issue of ordinariness. On the contrary, it is possible to find incompatible solutions. A solution to the relevance issue might render existing solutions to information retrieval ineffective or in need of revision. Similarly, adopting a particular solution to the ordinariness issue might preclude any hope of addressing the relevance problem.[20] All of this must be taken into account before deeming a potential answer as successful.

This worry also goes for emotions. They have potential roles to play in both issues: how cognitive knowledge is stored and how it can be exploited in a relevance-sensitive way. Therefore, care must be taken not to conflate these issues. Falling into this trap is easier than it may seem. Trying to account for relevance-sensitivity by developing methods for storing cognitive knowledge is practically the whole story of attempted solutions to the frame problem *qua* relevance issue. That's why, in order to concentrate on our particular topic, we'll explicitly put the problem of ordinariness aside and concentrate on the frame problem as a challenge regarding relevance-sensitiveness.

2. The possible roles of emotions in cognition

The claim that emotions provide a solution to the frame problem is rather vague. However, developing and discussing a very detailed account would narrow the scope of the discussion too much. Any such account

[17] For two very distinct suggestions, see SHANAHAN, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia,* and BARTH, C. *Representational cognitive pluralism: towards a cognitive-science of relevance-sensitivity*. PhD thesis — Belo Horizonte: Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, 2024.
[18] CHURCHLAND, P. M. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: The MIT Press, 1989.
[19] CLARK, A. Local Associations and Global Reason: Fodor's Frame Problem and Second-Order Search. *Cognitive science quarterly*, n. 2 (2002) p. 115-140.
[20] As an instance of this point, see BARTH, *Representational cognitive pluralism: towards a cognitive-science of relevance-sensitivity*. There I argue that relying on LOT (language of though) renders the frame problem unsolvable.

would inevitably assume a cognitive framework, a description of how the frame problem emerges within that framework, a conception of what emotions are, and so on. Thus, we need to find a middle ground. Rather than evaluating each account individually, we need a framework in which they can be formulated and assessed. This framework should capture their common elements, enabling us to concentrate on the set of possible roles that emotions can play within cognition. The question then becomes: given the possible roles of emotion within cognition, could any of these be the key to solving the frame problem?

We can find a proper candidate in Linquist and Bartol.[21] They dub it the *somatic marker model* (SMM). The idea of somatic markers is familiar to many researchers, especially in the form of the hypothesis presented by Damasio.[22] Roughly, a somatic marker is the way in which an emotion interacts with cognitive processing. An example is the feeling of anxiety or physical discomfort that a person feels when considering a risky decision, such as investing all their savings in an uncertain business. This negative feeling serves as a bodily signal (somatic marker) that helps the person evaluate possible negative consequences. At this level of description, somatic markers seem compatible with the aforementioned segregative stance. Indeed, they can amount to a piece of information that's available for processing, but that's not necessarily the case. A somatic marker can also be regarded as a physiological marker, i.e. something that constrains reasoning "from the outside". In this sense, they amount to a general organismic tendency to prefer or disfavor certain reasoning trajectories.

What matters to our current purposes, though, is that somatic markers may play multiple roles within cognition. Even though Damasio's claims clearly involve distinct roles, he never makes these distinctions explicit. Thus, SMM can be regarded as an effort to clarify the different cognitive roles for emotions that are identified in the works of Damasio and others.

In SMM, the decision-making process is broken down into five stages. The first is when the system realizes the need to make a decision about something. A change in the current circumstances might alert the system that something needs to be done, i.e. a new context has emerged. The second stage comprises formulating candidate options. In many contexts, the possible choices get pinned down from the outside: if coffee and tea are the only available beverages, then those are the only two options. But in other contexts, it is up to the system to populate the set of acceptable alternatives.

---

[21] LINQUIST, S.; BARTOL, J. Two Myths about Somatic Markers. *The British Journal for the Philosophy of Science*, v. 64, n. 3 (2013), p. 455-484.
[22] DAMASIO, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books, 1994.

The third stage consists of deliberation. The system must identify the set of relevant implications for each candidate choice. Identifying what's relevant is essential to decide when to stop deliberating, i.e. how to adaptively limit one's investment in time and energy. In the fourth stage, the system assigns value to the surviving alternatives. It already has a set of acceptable choices and has already determined their relevant implications. Now is the time to pick out the "winner", i.e. to effectively select one of them. At this point, a somatic marker connected to a given choice might become a kind of "vote" for it. Notice that such a marker need not be specifically connected to a single choice. Rather, it might be a vote for a particular articulation of distinct possibilities: it won't count as a vote for neither A nor B, but it may influence towards some articulation such as "first A and then B".

Finally, the fifth stage is where somatic markers may participate in the system's motivation to engage in the chosen course of action. The idea is that somatic markers would help avoid a kind of apathy, where the agent can determine the most appropriate course of action but doesn't care enough to effectively engage in it.

SMM's stages do not necessarily amount to clear-cut time periods, i.e. the idea is not to think of decision making as essentially serial. Rather, the goal is to make room for distinct contributions of somatic markers. A fully developed hypothesis can even conflate some of them. For instance, there could be an account where stages 2 and 3 are combined and managed by a single type of mechanism or process. The point of SMM is to provide an empirical dimension over which hypotheses like these can be articulated.

Another advantage of SMM is that it highlights the fact that there can be clear roles for emotions in cognition that are not connected to the role of emotions in solving the frame problem. Emotions can be involved in stage 1 inasmuch as they enable the system to take a situation as problematic, i.e. as something that requires engagement. Likewise, they can be involved in stage 5 and increase the chances that the system effectively engages in whatever course of action is required. These roles comprise what Haugeland famously characterized as the most crucial gap between humans and AI systems: computers don't give a damn.[23] Thus, the claim that emotions play a fundamental role in cognition is compatible with the claim that emotions have no place in solutions to the frame problem.

As an example of SMM's usefulness, consider the work of Evans. His view can be summarized through the following quote:

> Emotions prevent us from getting lost in endless explorations of potentially infinite search spaces by providing us with both the right kind of test and

[23] HAUGELAND, J. Understanding Natural Language. Em: *Having thought*. Cambridge: Harvard University Press, 1998, p. 47-60.

> the right kind of search strategy for each kind of problem we must solve. Only when emotions fail us do we end up in Hamlet's situation, suffering from a severe case of analysis paralysis.[24]

This is what he refers to as the *search hypothesis of emotion*, which posits a non-rational, emotion-governed process that establishes the boundaries of the information one considers. It does this by highlighting only the elements relevant to our needs and goals. In other words, emotions help delimit the range of possible consequences in rational decision-making processes.

Of course, as it stands, this falls short of a full-fledged solution. It indicates that emotions enable a system to avoid getting lost in a sea of irrelevant possibilities. But these emotions are introduced as a kind of black box that somehow provides incredibly appropriate answers to cognitive mechanisms whenever they are uncertain about pursuing a given line of reasoning.

Evans is aware of this, as he claims that the plausibility of his suggestion is deeply connected to the plausibility of an account of what emotions are. Thus, Evans' primary point seems to be that emotions should not be regarded as enemies of reason. Quite the opposite: what we call rationality is simply a nickname for emotion-constrained reasoning. After all, an agent spending hours making all sorts of valid yet utterly irrelevant claims would hardly be regarded as rational.

Evans' vocabulary aligns with that found in classic AI and classic cognitivism, particularly the approach that conceives of cognitive processing as searching through possible paths toward a solution.[25] In this view, decision-making involves exploring the space of possibilities using some search strategy (somewhat like finding the best path toward a goal in an incredibly complex map). Importantly, this does not necessarily commit Evans to any claim about the architecture of the human mind. Rather, it only assumes that, at some level of abstraction, decision-making can be understood in these terms. This holds true even if the underlying human wetware operates with rather distinct mechanisms.

The example of Evans shows that SMM helps us more clearly delineate the boundaries of the interplay between emotions and the frame problem. If one thinks that the frame problem can be isolated as an issue within stage 3, or even within a cluster of stages such as 2-3-4, that will surely affect what one regards as a plausible solution. Evans' suggestion is clear

---

[24] EVANS, D. The Search Hypothesis of Emotion. *The British Journal for the Philosophy of Science*, v. 53, n. 4 (2002), p. 497-509, aqui p. 503.
[25] NEWELL, H. A., Allen; Simon. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, v. 19 (1976), 1 mar. 1976.

in that it does not account for the role of emotions in any stage other than 3, even though one could easily extend it.[26]

The question of the scope of the frame problem brings another issue to the fore: what would it take for something to solve it? Evans claims that unless we understand what an emotion is, the question is bound to remain open. That is, perhaps the best we can do to tackle the frame problem is to develop the best theory of emotion we can, and then examine its possible influence on rational deliberation.

But is that really so? Like any unsolved investigation, even if we can't pinpoint who's guilty, we may have enough information to rule out a bunch of persons of interest. Similarly, even though the phenomenon of emotion is rich and multi-faceted, and we might be talking about multiple different kinds of states or processes, we can still discuss what it would take for them to play the required role.

3. What would it take for something to count as a solution?

Ransom faces the same question we're tackling here: what can emotions do about the frame problem? [27] Her view is rather pessimistic: not much, if anything. What is interesting about her position, however, is why she does not adopt a complete pessimism regarding the possibility of emotions playing a non-trivial role in a solution.

Ransom relies on the analysis of the frame problem provided by Wheeler.[28] Wheeler sees it as a two-headed beast. The first head comprises what he dubs the *intra-context frame problem*. It is about realizing what's relevant within an already established context. The second head is that of the *inter--context frame problem.* It is about realizing what's relevant to determine the context one is in, to begin with.

What can Wheeler's analysis buy us? It seems to offer this insight: while handling intra-context issues, the system can take the involved domains

[26] This is not a criticism of Evans's stance. Evans may surely think (or at least be open to the idea) that the frame problem encompasses other stages; the point here is just that in the treatment presented in *The Search hypothesis of emotion*, that is not the case. Indeed, Evans makes a clear effort to dissociate from discussions about what the frame problem is. One of the goals of this paper, however, is to show that trying to disassemble the frame problem into sets of narrower problems is a non-starter. This affects even those who end up meeting one of these narrower problems under another name, and try to handle it in isolation, unaware that there's a huge beast underneath.

[27] RANSOM, Why Emotions Do Not Solve the Frame Problem.

[28] WHEELER, Cognition in context: phenomenology, situated robotics and the frame problem.

for granted. This allows it to apply cognitive resources suitable for those domains, such as domain-specific heuristics, tricks, and shortcuts. All of this makes the process computationally tractable, and the system's ability to focus on what is circumstantially relevant is no longer mysterious. In contrast, the inter-context head is harder to tame. Establishing the appropriate context of operation draws on nearly every cognitive resource the system can integrate, and this quickly becomes unmanageable.[29] Therefore, given Wheeler's analysis, even if we can't figure out how the system can determine its current context of operation, we can at least take cases in which this was already done and work on them in isolation.

Ransom's position is that emotions can't solve either of these problems on their own. However, though they can't do a thing for the inter-context dimension, they *can* be of help at least with the intra-context dimension.

Let us first examine why she believes that emotions offer no solution to the inter-context dimension. Consider what happens at stage 1 of SMM: recognizing a circumstance that calls for action or deliberation. This comprises understanding the context one is in (or should consider itself to be in). If somatic markers act as labels for the circumstantially relevant features for context selection, the claim that they completely solve the inter-context problem would imply that they label all and only the contextually relevant features.

However, even though we lack a full-fledged theory of emotions, we do know that emotions are themselves context-dependent. The same input might be satisfying or disturbing, depending on the circumstances in which one encounters it. This means that, before context selection, the effects of any somatic markers are undefined or at best under-determined. It should go without saying that if the effects of somatic markers cannot be determined before context determination, they won't be helpful in context selection tasks.

Ransom is more optimistic about the potential role of emotions in the intra-context version. While somatic markers cannot provide a full solution, she accepts that, within an already established context, emotions can provide valence information, which *"may help to solve the intra-context frame problem"*.[30] Somatic markers may establish the threshold for how negative a course of action must be to be avoided, as well as indicate the current level of affective positivity or negativity of some options. In SMM terms, provided you're already thinking within a context that was somehow established at stage 1, you're better off at stages 2, 3 and 4.

---

[29] Importantly, this integration need not be full-fledged or isotropic in Fodor's sense. All that is required is that the set of possible articulations is so large that exhaustive consideration of every possibility becomes computationally intractable.

[30] RANSOM, Why Emotions Do Not Solve the Frame Problem, p. 14.

That makes sense because, if one's already within a given context, one can assume that choices outside of that context have already been ruled out. Thus, the space of possibilities is significantly narrowed down, leaving only a relatively small set of remaining choices. Heuristic approaches may then be sufficient. In particular, Evans' approach might be reinterpreted as having the intra-context frame problem in view.

In despite of that, Ransom acknowledges many limitations to the possible role that emotions may play within contexts. They are not the only significant factor, and there is no guarantee that every possibility will have a somatic marker attached, especially in more offline, theoretical thinking. There is no direct connection between being relevant and being marked. Thus, even though somatic markers might be associated with possible outcomes or courses of action, their role is certainly not decisive.

In a nutshell, the reason why emotions seem helpful within well-defined contexts is the same reason why they can't help when determining the context: they are context-dependent. This can be generalized, as any context-dependent element presents the same virtues and limitations as emotions. Being context-*independent* is, therefore, a clear requirement for any potential candidate to be part of a solution to the inter-context frame problem. However, the potential role for emotions within contexts that Ransom admits relies on the claim that it makes sense to talk of context--bounded relevance determination. But is that so?

4. There is no context-bounded relevance determination

I think Ransom's reasoning is flawless, provided we accept Wheeler's analysis of the frame problem as bi-dimensional. I don't think we should accept Wheeler's analysis, though.

Wheeler's account is an instance of what I'll call the *subset approach*. Broadly, it conceives a context as a specific configuration of a subset of cognitive resources.[31] The central idea is that the relevance of a feature is modeled by its presence within a given structure. Conversely, the absence of an element from the structure would indicate its irrelevance. Thus, a system guided by any such context can just assume that every feature present is relevant, and everything else can be safely disregarded as irrelevant.

[31] The term "cognitive resources" is indented to include both internal and external states that bear on the system's behavior.

This is the same strategy found in the earliest attempts to solve the frame problem within GOFAI.[32] The data structures of Minsky[33] (confusingly called frames), which inspired the scripts of Schank and Abelson, share the same core idea.[34] In Schank's famous restaurant example, the goal was to list the relevant features for typical behavior (tables, menu, beverages, etc.) and their usual sequence (first you ask for a table, then you ask for a menu, and so on). Wheeler is not advocating that contexts should be modeled as data structures, but the whole point of distinguishing intra--context and inter-context issues is the possibility of handling them in isolation, and that requires delineating a subset of the system's resources.[35] In this sense, within any already established context, the processing can be encapsulated and safely ignore whatever is left out.

How plausible is this kind of encapsulated reasoning? We can surely conceive the existence of local mechanisms that are largely insensitive to the circumstances, but if our goal is to explain human's flexibility, that can't be the rule.[36] That's because, even though we might be, at any given moment, working within the boundaries of an already established context, we don't loose sight of how extra-context features can become relevant. For instance, our behavior in restaurants is sensitive to an open-ended set of features that, despite being relevant, are not directly related to what's usually the case within restaurants. The absence of work colleagues can be crucial while deliberating about what (or how much) to drink. Likewise, the presence of a work colleague with a history of alcoholism can make us retreat from suggesting getting a beer.

Consider also the question whether it is plausible to leave a restaurant because one's favorite dish is unavailable. The answer relies heavily on how one characterizes the context. If one is having birthday-dinner with friends and gets a gift characterized as "paying the bill", then it seems a

---

[32] *Good old-fashioned AI*. The name comes from HAUGELAND, *Artificial Intelligence: the very idea.*
[33] MINSKY, M. A framework for representing knowledge. Em: HAUGELAND, John (Ed.). *Mind design II: phylosophy, psychology, artificial intelligence.* Cambridge, MA: MIT Press, 1997, p. 111-142.
[34] SCHANK, R. C.; ABELSON, R. P. *Scripts, Plans, Goals and Understanding.* London: Taylor & Francis Inc, 1977.
[35] I'm not claiming that there's no difference between Wheeler's and Schank's approaches. Notice, for instance, that Schank tries to handle the frame problem by articulating a way to store information, i.e. by modeling ordinariness. As for Wheeler, he is silent on this matter.
[36] Perhaps some friends of massive modularity might insist that local mechanism *can* be the rule. As Sperber shows in SPERBER, D. Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context-Sensitive. Em: CARRUTHERS, Peter; LAURENCE, Stephen; STICH, Stephen (Eds.). *The Innate Mind: Structure and Contents*. Oxford: Oxford University Press, 2005, p. 53-68, massively modular systems can be surprisingly flexible. However, that doesn't really buy what's needed to solve the frame problem, for now it comprises the challenge of explaining how the system manages to articulate its huge set of local mechanisms in a way that's sensitive to contextual relevance. See BARTH, *Representational cognitive pluralism: towards a cognitive-science of relevance-sensitivity*.

bit too much to make everybody leave to another restaurant just because one's favorite dish is unavailable. But if one's gift is characterized as "getting his/her favorite dish", then the option will be much more salient and acceptable for everybody involved.

The crucial point is that any attempt to model feature irrelevance as feature absence will result in a model that underdetermines the context in which the action (or the thinking) is taking place. To avoid this, one would have to predict in advance all the possible features that might arise within that context, and this inevitably involves considering extra-context features. After all, no one expects to find the need to fully characterize the gift one's being given on that day as an essential part of contexts such as being-on-a-restaurant or having-dinner-with-friends. In this picture, the question becomes: at which point is the context's set of relevant features fully determined?

If we leave the determination of every single relevant feature of the full--fledged context to the stage of context determination, then Wheeler's analysis collapses. There is no intra-context relevance determination, only intra-context normal processing. The frame problem would have already been solved at the context determination stage, meaning the inter-context stage carries all of the burden. Since emotions play no role in context selection, this line of reasoning implies that there's no place for emotions in any solution to the frame problem.

On the other hand, if we leave the determination of relevance to happen also within the context, as per Wheeler's analysis, we can't explain how we can be sensitive to features outside the context's boundaries. The whole idea was to break down the frame problem into two distinct issues. First, the system selects the relevant context, and then the system selects the relevant features based on the subset of features comprising that context. But if one is required to continuously review what is in and out of the context, then once again the burden is entirely carried by the inter-context dimension.

In order to save his analysis of the frame problem as bi-dimensional, Wheeler must rely on a wanting theory of context determination. Such a theory would need to provide principled grounds for determining what's in and out a given context. It provides types or stereotypes of contexts, if you like. However, as the previous examples have shown, proper behavior within restaurants involves information about lots of distinct domains: human health, work colleagues, and even *minutiae* such as the nature of a gift being given on the occasion. The challenge, then, is to find a middle ground between: 1) a context-determination process that does not dilute into selecting everything the system is capable of attending to, which would defeat the purpose of having a context; and 2) a context-determination

process that leaves too much out and fails to enable context-bounded relevance determination.

Wheeler suffered similar pressure from Dreyfus[37] and Rietveld.[38] Both claim that the real work regarding relevance determination is done solely at the inter-context level. Intra-context processing assumes, rather than provides, a solution to the frame problem. In his reply, Wheeler admits that the idea of a intra-context frame problem depends on the possibility of a theory of context determination:

> This is, I admit, a genuine risk, one that casts a spotlight on the Herculean theoretical challenge of giving a theory of how our cognitive systems determine where the boundaries of contexts lie. It seems to me, however, that at this juncture we confront a choice. Either we remain committed to the distinction between intra-context and inter-context sensitivity to relevance, and so humbly accept the challenge to give a theory of context determination, or (…) confront the fact that (…) our (…) cognitive science of relevance-sensitivity in general, and not just of context-switching, has barely begun.[39]

Wheeler seems to believe that we should stick to the notion of intra--context and assume that there must be some acceptable theory to ground it. Otherwise, we would have to assume that there's never been any real progress towards a solution to the frame problem. Well, I have serious doubts about the existence of such a theory. It is worth noticing that Wheeler's challenge is very similar (if not identical) to the one faced by Schank's scripts decades ago. Perhaps it's time to admit that this is a flawed challenge and reconsider the way we're posing the question.

Ultimately, I think Wheeler's insistence is a product of conceiving contexts as subsets of resources. It may be more fruitful to regard contexts as a "tonality" that the cognitive world of the agent takes. A more or less global state in which some things are considered highly relevant for that particular situation, and other things are further away, but never completely fade. Thus, when it comes to context-switching, rather than picking out different portions of cognitive resources in the form of stereotypes (or something similar), one ought to think of them as a trajectory through the (more os less) global state space of the system. Contexts are not subsets of the system's cognitive resources, but rather more or less global articulations

---

[37] DREYFUS, H. L. Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. In: *Artificial Intelligence*. Elsevier, v. 171 (2007), p. 1137-1160.
[38] RIETVELD, E. Context-switching and responsiveness to real relevance. Em: KIVERSTEIN, Julian; WHEELER, Michael (Eds.). *Heidegger and cognitive science*. Hampshire: Palgrave Macmillan, 2012. p. 105-135.
[39] WHEELER, M. Naturalizing Dasein and other (alleged) heresies. Em: KIVERSTEIN, Julian; WHEELER, Michael (Eds.). *Heidegger and cognitive science.* Hampshire: Palgrave Macmillan, 2012. p. 176-212, aqui p. 209.

of the system's overall set of capabilities. This view has the advantage that we are not tempted to mistake a redescription of the frame problem for real progress, as we might when pursuing context-bounded reasoning.

Such a mistake has been made over and over throughout the history of solution attempts. As an example of this point, let us consider again the idea of (existential) moods.[40] Being depressed, being in love or in peace can be understood as partial characterizations of our stance towards the world. Thus, one could grant the point that I've been doing here (emotions have no key role in solutions to the frame problem), and still think that moods have a role to play, at least potentially. Couldn't moods comprise a way to "frame" the world? Couldn't they be able to constrain, at least partially, the set of possible things to consider while thinking, i.e. at least partially determine the current context? Such possibility is appealing, for moods are not context-sensitive, at least not in the way emotions are. Maybe moods can help us understand what grounds one's trajectory from a contextual tonality to another?

I think moods do constrain our thinking. However, the frame problem is not about finding *some* way to constrain our reasoning. It challenges us to find a way to do that without losing track of what's relevant in an ever-changing and open-ended set of contexts. That's why, as it stands, the reliance on moods strikes me as just another way to mistake a redescription of the issue for real progress. Moods can partially constrain our thinking and our behavior, but so can Schank's scripts. And as we have seen, the fact that scripts can afford only partial constraints (i.e. they can only underdetermine contexts), is the reason they can't buy us real progress.

Well, one might ask: couldn't we rely on both moods and scripts? Or maybe add up a plurality of such context-independent sources? As I see it, that only seems plausible if we let the subset approach comes back through the window. If we think of contexts as subsets of features, the more determinate features we have, the closer we feel to a solution. But, as discussed, human contexts are always subject to an open-ended set of features. Consider what would happen if we attempt to characterize a context by relying on (say) both scripts and moods. We could articulate a mood like "being in love" and a script such as "restaurant", resulting in something like "being in love in a restaurant". What would follow from that? Which features of one's environment are supposed to be made more (or less) salient? The problem of how to articulate scripts in order to handle relevance is now the problem of how to articulate scripts *and moods* in order to handle relevance. The open-ended character of the contexts in which we carry on our tasks makes all the difference. If one takes an open-ended set of elements that

[40] I'm deeply thankful to Felipe Carvalho for his thoughtful comments and for stressing the need to address this.

must be properly articulated and slices it in two, what one gets are two open-ended sets that must be properly articulated. Therefore, the problem remains the same: how are we to articulate the involved resources, be they scripts, moods or whatever?[41] At this point, it should be clear that we're not making progress. We're just redescribing the issue.

All of this emphasizes what's peculiarly challenging about the frame problem: its striking resistance to being analyzed into smaller, simpler problems and its consequent resilience against gradual progress. That's the difficulty faced by Wheeler's. Thus, by rejecting Wheeler's analysis, we must also reject Ransom's claim that emotions can help address the frame problem. Such a claim could only make sense within the intra-context dimension, but in the picture we've drawn here, there's no such distinction.

Conclusion

At first glance, emotions seem to offer a promising approach to finally making progress towards a solution to the frame problem. Emotions are ubiquitous in our cognitive lives and deeply connected to our needs and goals. What better candidate to handle a similarly ubiquitous capacity, such as determining what's relevant in an open-ended set of contexts?

Unfortunately, even though there is still much to learn about the role of feelings in general (and emotions in particular) within cognition, what we know so far already compels us to put the brakes on solving the frame problem through emotions. Even if emotions have significant roles to play at every level of SMM, none of these roles buy us anything, inasmuch as frame problem is concerned. And we know why: relevance is deeply context-dependent, and this precludes the possibility that any context--dependent feature can be part of a solution.

As a result, there seems to be no place for emotions in solutions to the frame problem. It is important to be clear, though: I am not claiming that an eventual solution to the frame problem will not assume a particular way to store cognitive knowledge or the existence of somatic markers. Rather, the point is that the frame problem is about how our cognitive resources are articulated and exploited, whatever they are (i.e. whether they are affective or not). To play a role in a solution to the frame problem is to play a role in this articulation. Up to this point, whenever we tried to come up with

[41] One can evidently insist that felt qualities, emotions and moods are not exhaustive of the affective dimension and claim that something within that dimension can still play some yet-to-be-specified crucial role in a solution. But determining whether something within the dimension of affectivity can play a cognitive role distinct from those considered here is not the burden of this work.

strategies to articulate cognitive resources in a relevance-sensitive way, all we could do was finding more kinds of resources in need of relevance-sensitive articulation. Something is still missing, and whatever is missing carries the burden of handling (i.e. being a solution to), the frame problem.

References

BARTH, C. *Representational cognitive pluralism: towards a cognitive-science of relevance-sensitivity*. PhD thesis — Belo Horizonte: Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, 2024.

CHURCHLAND, P. M. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: The MIT Press, 1989.

CLARK, A. Local Associations and Global Reason: Fodor's Frame Problem and Second-Order Search. *Cognitive science quarterly*, n. 2 (2002), p. 115-140.

DAMASIO, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books, 1994.

DENNETT, D. Inteligência artificial como filosofia e como psicologia. Em: *Brainstorms: ensaios filosóficos sobre a mente e a psicologia*. Tradução de Luiz Henrique De Araújo Dutra. São Paulo: Unesp, 1999[1978]. p. 163-183.

DREYFUS, H. L. Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Em: *Artificial Intelligence*. Elsevier, v. 171 (2007), p. 1137-1160.

EVANS, D. The Search Hypothesis of Emotion. *The British Journal for the Philosophy of Science*, v. 53, n. 4 (2002), p. 497-509.

FODOR, J. A. Modules, frames, fridgeons, sleeping dogs and the music of the spheres. Em: PYLYSHYN, Zenon W. (Ed.). *The robot's dillemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex, 1987, p. 139-149.

FORD, K. M.; PYLYSHYN, Z. W. (EDS.). *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*. Norwood, NJ, USA: Ablex Publishing Corp., 1996.

HAUGELAND, J. *Artificial Intelligence: The Very Idea*. Cambridge, MA: The MIT Press, 1985.

HAUGELAND, J. An overview of the frame problem. Em: PYLYSHYN, Zenon W. (Ed.). *The robot's dillemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex, 1987, p. 76-93.

HAUGELAND, J. Understanding Natural Language. Em: *Having thought*. Cambridge: Harvard University Press, 1998, p. 47-60.

HAYES, P. J. What the frame problem is and isn't. Em: PYLYSHYN, Zenon W. (Ed.). *The robot's dillemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex, 1987, p. 123-137.

LINQUIST, S.; BARTOL, J. Two Myths about Somatic Markers. *The British Journal for the Philosophy of Science*, v. 64, n. 3 (2013), p. 455-484.

MCCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, v. 4 (1969), p. 463-502.

MESQUITA, B.; BOIGER, M. Emotions in Context: A Sociodynamic Model of Emotions. *Emotion Review*, v. 6, n. 4 (2014), p. 298-302.

MINSKY, M. A framework for representing knowledge. Em: HAUGELAND, John (Ed.). *Mind design II: phylosophy, psychology, artificial intelligence*. Cambridge, MA: The MIT Press, 1997, p. 111-142.

MINSKY, M. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006.

NEWELL, H. A., Allen; Simon. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, v. 19, 1 mar. 1976.

PYLYSHYN, Z. W. (ED.). *The Robots Dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex, 1987.

RANSOM, M. Why Emotions Do Not Solve the Frame Problem. Em: MULLER, Vincent (Ed.). *Fundamental Issues of Artificial Intelligence*. Berlin: Springer, 2016, p. 353-365.

RIETVELD, E. Context-switching and responsiveness to real relevance. Em: KIVERSTEIN, Julian; WHEELER, Michael (Eds.). *Heidegger and cognitive science*. Hampshire: Palgrave Macmillan, 2012, p. 105-135.

SCHANK, R. C.; ABELSON, R. P. *Scripts, Plans, Goals and Understanding*. London: Taylor & Francis Inc, 1977.

SHANAHAN, M. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. Cambridge, MA: The MIT Press, 1997.

SOUSA, R. *The Rationality of Emotion*. Cambridge: MIT Press, 1987.

SPERBER, D. Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context-Sensitive. Em: CARRUTHERS, Peter; LAURENCE, Stephen; STICH, Stephen (Eds.). *The Innate Mind: Structure and Contents*. Oxford: Oxford University Press, 2005, p. 53-68.

WHEELER, M. Cognition in context: phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, v. 16, n. 3 (2008), p. 323-349.

WHEELER, M. Naturalizing Dasein and other (alleged) heresies. Em: KIVERSTEIN, Julian; WHEELER, Michael (Eds.). *Heidegger and cognitive science*. Hampshire: Palgrave Macmillan, 2012, p. 176-212.

Carlos Barth

Endereço profissional:
Faculdade Jesuíta de Filosofia e Teologia (FAJE)
Av. Dr. Cristiano Guimarães, 2127 – Planalto
31720-300 – Belo Horizonte