

Cognitive-Affective and Dynamic Modeling of a Sacrificial Decision: Integration of Emotional Appraisal, Cognitive Task Analysis, and Fuzzy Cognitive Maps

Cecilia Jokebed Rea-Castañares¹, Ana Lilia Laureano-Cruces^{2*}, Martha Mora-Torres²,
Ismael Martínez-Bonilla³

¹Maestría en Optimización, Universidad Autónoma Metropolitana-Azcapotzalco, CDMX, México

²Departamento de Sistemas, Universidad Autónoma Metropolitana-Azcapotzalco, CDMX, México

³Laboratorio de Educación y Evaluación Digital, Facultad de Estudios Superiores Iztacala, UNAM, CDMX, México

Email: *clc@azc.uam.mx

How to cite this paper: Rea-Castañares, C.J., Laureano-Cruces, A.L., Mora-Torres, M. and Martínez-Bonilla, I. (2026) Cognitive-Affective and Dynamic Modeling of a Sacrificial Decision: Integration of Emotional Appraisal, Cognitive Task Analysis, and Fuzzy Cognitive Maps. *International Journal of Intelligence Science*, 16, 234-247. <https://doi.org/10.4236/ijis.2026.163012>

Received: June 11, 2026

Accepted: July 13, 2026

Published: July 16, 2026

Copyright © 2026 by author(s) and

Scientific Research Publishing Inc.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Sacrificial decisions in contexts of high moral conflict are often analyzed through normative frameworks or experimental studies focused on sacrificial dilemmas. However, these approaches do not always allow representing the internal dynamics that connect environmental conditions, causal inferences, moral goals, emotions, and observable action. This article presents a cognitive-affective and dynamic model to analyze voluntary sacrificial behavior in a narrative case: the decision of an agent to surrender to an antagonist despite knowing that they will likely die, under the premise that their death may reduce collective harm. The objective was to construct and evaluate an integrative scheme combining cognitive task analysis as a technique for identifying elements to represent in the behavioral model; affective-moral appraisal structure as a cognitive representation of emotions in the modeled behavior; and finally, fuzzy cognitive maps as a technique for representing the behavioral model knowledge for computational implementation. The methodology included four phases: delimitation of the behavior and decision context; formalization of the mental model into facts, concepts, inferences, values, and action; construction of a causal matrix of seven nodes; and iterative simulation of scenarios with a binary activation function. The results show that in scenarios where active threat, ongoing war, and recognition of an internal causal condition coincide, the system converges toward a stable state composed of three nodes: minimizing deaths, voluntary sacrifice, and weakening the enemy. The discussion proposes that the behavior can be interpreted as an emergent prop-

erty of the system, not as an impulsive response nor as a purely utilitarian decision. It is concluded that fuzzy cognitive maps offer a useful formal strategy to represent complex moral decision processes, provided their results are interpreted as modeling hypotheses and not as direct empirical validation of human behavior. A corrective modeling refinement is also specified to address possible overactivation of the sacrifice node: the revised simulation distinguishes necessary from sufficient conditions by incorporating a thresholded sigmoidal activation rule for voluntary sacrifice and treating the internal causal condition as an enabling condition rather than as an isolated sufficient cause.

Keywords

Cognitive Engineering, Fuzzy Cognitive Maps, Moral Decision Making, Emotional Appraisal, Cognitive Task Analysis, Computational Modeling

1. Introduction

The analysis of complex moral decisions requires going beyond the description of behavior. In situations of collective harm, persistent threat, and extreme individual cost, observable behavior usually constitutes the last link in a chain of representation, appraisal, and inference. From cognitive science and cognitive engineering perspectives, human decision-making in uncertain environments can be understood as the result of how an agent perceives the situation, anticipates consequences, prioritizes goals, and translates an action criterion into an executable response. In this sense, sacrificial decisions constitute a case of special interest because they require modeling the tension between self-preservation, moral responsibility, and collective benefit [1].

Contemporary literature on moral judgment has pointed out that moral decisions cannot be reduced to abstract rules or linear consequence calculations. Greene [2] argues that contemporary moral cognition has shifted from exclusively philosophical-normative models toward psychological explanations integrating emotion, reasoning, conflict, and control. Complementarily, Crockett [3] suggests that formal models can help clarify the mechanisms by which people articulate preferences, values, and choices in moral dilemmas. From this perspective, the problem is not limited to determining which action should be considered correct, but to explaining under which internal and external conditions an action becomes selectable for an agent.

Likewise, various studies have shown that moral judgment can be modified depending on task realism, bodily proximity, attributed responsibility, and the transition between hypothetical judgment and simulated action [4] [5]. This difference between deciding what would be morally acceptable and executing moral conduct is central to delimiting the scope.

Interest in these problems has also extended to the field of computational ethics

and intelligent system design. Studies on autonomous vehicles have shown that dilemmas of inevitable harm are not only philosophical problems but also modeling, decision, and design problems under constraints [6] [7]. However, much of this literature has focused on normative preferences or aggregated participant responses, while less attention has been paid to formalizing the internal architecture of a particular moral decision. This gap justifies the development of models that allow representing, within a single system, environmental conditions, causal inferences, goals, emotions, values, and action.

In this context, cognitive appraisal theories (OCC) offer a relevant framework to explain how events acquire affective meaning in relation to the agent's goals, expectations, and standards. Emotion is not understood here as interference opposed to reason but as an evaluative signal that organizes action according to its relevance, urgency, and expected implications [8] [9]. Similarly, Scherer and Moors [10] argue that emotional processes involve differentiation of components and evaluation of events with respect to goals, coping possibilities, and normative meaning. This approach allows analyzing emotions such as concern, empathy, hope, or relief not as isolated causes of behavior but as appraisal markers that modify the strength of certain decision trajectories.

From this perspective, a sacrificial situation can activate simultaneously negative and positive appraisals. The threat to individual survival may generate sadness or concern, while the possibility of reducing collective harm may activate hope or relief. Consequently, extreme moral behavior should not be interpreted as a purely rational action nor as an immediate affective reaction, but as an integration among prioritized goals, causal inferences, and affective-moral appraisal [11]-[14]. This approach avoids two frequent reductionisms: considering emotion as irrational noise or reducing moral decision to an affective impulse.

However, a limitation of emotional appraisal models is that they often describe categorical or functional relationships without explicitly representing the system's temporal evolution. To address this limitation, fuzzy cognitive maps (FCMs) offer a causal-dynamic modeling strategy. Since Kosko's original formulation [15], fuzzy cognitive maps allow representing concepts as nodes connected by weighted causal relationships. Iterative propagation of activation allows observing how a system changes over time, how certain configurations stabilize, and how feedback modifies initial conditions. In recent years, FCMs have been used as flexible tools to represent expert knowledge, model complex systems, and analyze scenarios where qualitative and quantitative elements converge [16] [17].

From the above arises the specific problem addressed in this article: how to integratively model an extreme moral decision in which an agent chooses personal sacrifice as a strategy to produce collective benefit. Literature on moral judgment has shown that ethical decisions are neither exclusively rational nor exclusively affective. Classical and contemporary studies have indicated that emotion, cognition, and executive control jointly, though not always harmoniously, participate in moral judgment [18]-[22]. This evidence suggests that explanatory models

must represent not only the logical structure of the decision but also how affective evaluations contribute to prioritizing goals and legitimizing certain courses of action.

To address the above, a case was developed for this analysis, consisting of Harry Potter's decision to voluntarily go to the Forbidden Forest to surrender to Voldemort, even anticipating that he will likely die. The scene is considered here as a narrative case of high cognitive and moral density. The aim is not to study an empirical subject nor to perform an exhaustive literary interpretation of the work, but to use a narratively defined behavior as a controlled scenario to test a cognitive formalization procedure. The case is useful because it concentrates typical conditions of extreme decision: active threat, risk to third parties, internal causal restriction, action window, conflict between individual survival and collective welfare, and the need to translate a causal inference into immediate action [23]. The case is treated as a controlled cognitive-modeling scenario, not merely as an illustrative literary example, because the narrative explicitly specifies the agent's knowledge state, the causal constraint, the collective threat, and the action opportunity. Its value is therefore analytical transferability: the modeling procedure may be compared with non-fictional high-conflict decisions in which an agent assumes personal cost under perceived causal responsibility, without claiming statistical generalization from a fictional case.

The development of this analysis will allow articulating elements usually worked on separately: cognitive task analysis, emotional appraisal models, and fuzzy cognitive maps. Methodologically, it proposes a replicable route to transform complex behavior into an interpretable computational model. This can be relevant for analyzing human decisions in education, professional training, clinical simulation, intelligent agent design, tutoring systems, and learning environments where it is necessary to model not only what a person does but the structure of goals, affects, and constraints that make action possible.

The methodological novelty of this work lies not in using cognitive task analysis, OCC appraisal, or fuzzy cognitive maps separately, but in integrating them into a single operational pipeline: cognitive task analysis identifies the minimal decision components; OCC appraisal specifies the affective-moral valuation of events, actions, and outcomes; and the fuzzy cognitive map translates these components into a simulable causal-dynamic structure. This integration offers a replicable procedure for formalizing complex moral decisions, an approach that remains uncommon in prior conceptual applications of fuzzy cognitive maps and moral-cognition research [21].

Therefore, the following research question is posed: is it possible to represent an agent's sacrificial decision as an emergent behavior derived from the dynamic interaction among causal beliefs, moral goals, affects, and environmental conditions? From this question, the general objective of the study is to develop an affective-cognitive model of extreme moral decision-making through the integration of cognitive task analysis, genetic graph of behavior, OCC theory, and fuzzy

cognitive maps. The specific objectives were: *first*, to identify the minimal cognitive components that make the behavior rational within the narrative world; *second*, to organize these components into a structure of tasks, inferences, values, and actions; *third*, to formalize the affective and moral appraisal of events, actions, and objects; and *fourth*, to evaluate the system's dynamics through iterative simulations with different initial vectors.

2. Methodology

A conceptual computational modeling study was designed based on a narrative case. The study's approach and scope were exploratory-descriptive since cognitive and affective-moral units of the case were identified; subsequently, these units were translated into a causal structure suitable for simulation. No human participants, clinical data, or psychometric measurements were used. Accordingly, the study seeks analytical generalization of a modeling procedure, not empirical generalization about a population of decision makers. External validity is therefore addressed as transferability to structurally analogous cases, rather than as statistical representativeness.

2.1. Procedure

2.1.1. Case of Analysis and Target Behavior

The target behavior was defined as the decision of an agent to voluntarily move toward the place where the antagonist is located to surrender, despite knowing that they will likely die. The decision context included four minimal conditions: 1) active threat, because the antagonist retains capacity for action; 2) systemic harm, because the conflict's continuity implies present and future deaths; 3) internal causal restriction, operationally defined as the agent's recognition that their own existence functions as a necessary causal condition that partially sustains the antagonist's strength; and 4) action window, operationally defined as a limited situational opportunity in which the agent's action can still modify the causal trajectory of the conflict. These conditions delimit the internal rationality of the behavior: the action is not interpreted as a desire to die but as an extreme strategy to alter a causal network of harm. In the revised model, the internal causal restriction is treated as necessary for voluntary sacrifice, but not sufficient by itself; sacrifice requires its interaction with threat, expected collective harm, and moral self-regulation. The phases of behavior model development are described below.

2.1.2. Phase 1: Cognitive Task Analysis

The first phase consisted of decomposing the behavior into cognitive inputs, required skills, critical decisions, and plausible errors. External inputs were active threat, ongoing war, risk to third parties, and temporal opportunity to act. Internal inputs were self-knowledge of the causal condition, moral responsibility, and anticipated consequence evaluation. Identified cognitive skills were pattern recognition, prospective causal reasoning, moral self-regulation, and procedural planning. The functional sequence was synthesized into five operations: recognize, in-

fer, evaluate, decide, and act.

2.1.3. Phase 2: Mental Model Formalization

The second phase organized the agent's knowledge through an island-type structure using an F-C-I-A-S scheme (Factual/Facts, Conceptual, Inference/Cause, Values/Self-regulation, and Strategy/Action). This organization allows distinguishing information assumed as true, concepts structuring the problem, causal rules, values regulating choice, and final observable behavior. **Table 1** summarizes this decomposition. In this article, an island-type structure refers to the modular representation of qualitatively different knowledge blocks involved in the behavior; each island groups a type of knowledge that performs a distinct function in the decision architecture, while still remaining connected to the overall causal model.

Table 1. Agent knowledge blocks (Islands).

Level	Content	Function in the Model
F: Facts	The agent carries an internal causal condition; the antagonist remains active; the conflict continues.	Delimit the system's initial state and problem constraints.
C: Concepts	Necessary condition; weaken the enemy; individual cost versus collective benefit.	Transform facts into useful categories for reasoning.
I: Inferences	If the internal condition persists, there is no definitive victory; if eliminated, the antagonist weakens; if no action is taken, deaths increase.	Simulate consequences of acting and not acting.
A: Self-regulation	Priority to minimize others' deaths; acceptance of personal sacrifice for collective benefit.	Prioritize goals and resolve conflict between self-preservation and collective welfare.
S: Strategy and Action	Personal sacrifice as strategy; voluntary surrender as action.	Convert the cognitive-moral structure into observable behavior.

2.1.4. Phase 3: Affective-Moral Appraisal Structure

The third phase translated events, goals, and moral standards into an appraisal structure. Active goals were defeating the enemy, minimize deaths, protect third parties, end the war, and survive. Individual survival remained a relevant goal but with lower priority compared to collective goals. Negative events (continuation of war and future deaths) were linked to concern and empathy; the sacrifice event was associated with hope due to its instrumental function; and the possibility of defeating the antagonist was associated with relief. Actions were evaluated according to standards: sacrificing to reduce collective harm was classified as morally

correct, while deliberately causing harm was classified as morally reprehensible. This structure informed the selection of nodes for the fuzzy cognitive map. To make the transition from appraisal to FCM explicit, each appraisal component was mapped either to a node or to a causal edge. Concern and empathy related to future deaths informed the links from conflict and anticipated harm toward C5; hope regarding the usefulness of sacrifice informed the connection between C5 and C6; anticipated relief regarding weakening the antagonist informed the reinforcement between C6, C7, and C5; and moral responsibility derived from the internal causal condition informed the enabling role of C1 in activating C6. The foregoing is detailed in **Table 2**.

Table 2. Translation of affective-moral appraisal into FCM components.

Appraisal input	Affective marker	FCM translation	Modeling rationale
Continuation of war and future deaths	Concern and empathy	$C3 \rightarrow C4$; $C4 \rightarrow C5$; $C5 \rightarrow C6$	Anticipated harm increases the priority of minimizing deaths and creates pressure toward action.
Internal causal restriction	Responsibility and negative appraisal of self-preservation	C1 as enabling condition for C6; $C1 \rightarrow C7$	Sacrifice is causally meaningful only if the agent's condition is connected to weakening the antagonist.
Possibility of weakening the antagonist	Hope and anticipated relief	$C6 \rightarrow C7$; $C7 \rightarrow C5$	Expected resolution reinforces both the strategic outcome and the moral goal.
Moral standard against avoidable collective harm	Moral obligation/self-regulation	C5 as regulatory node; $C5 \rightarrow C6$	The collective goal prioritizes action over self-preservation when the causal condition is present.

2.1.5. Phase 4: Construction of the Fuzzy Cognitive Map

Finally, seven nodes were defined: C1, the agent carries an internal causal condition; C2, active antagonist; C3, ongoing conflict; C4, future deaths; C5, minimize deaths; C6, voluntary sacrifice/surrender; and C7, weakening of the antagonist. The nodes were derived from the preceding cognitive task analysis and F-C-I-A-S decomposition rather than selected independently: facts informed C1-C4, self-regulatory values informed C5, strategy/action informed C6, and expected causal outcome informed C7. Causal weights were assigned through author-consensus analytical coding based on three rules: first, a positive value represents a directly proportional influence; second, a negative value represents an inversely proportional influence; and third, the magnitude of the value represents the relative intensity of the causal implication in the modeled scenario. Thus, the matrix should be interpreted as an explicit modeling hypothesis derived from the previous phases, not as an empirically estimated parameter matrix. **Table 3** presents the causal matrix of these nodes, including the revised interpretation of C6 as a thresholded and C1-enabled action node.

Table 3. Causal matrix with corrective threshold interpretation.

Origin/Destination	C1	C2	C3	C4	C5	C6	C7
C1	0	0	0	0	0	0.50	1
C2	0	0	1	1	0	0.25	-1
C3	0	0	0	1	1	0.25	0
C4	0	0	0	0	1	0.25	0
C5	0	-1	-1	-1	0	0.75	1
C6	0	-1	-1	-1	1	0	1
C7	0	-1	-1	-1	1	0	0

Corrective rule for C6: the values directed toward voluntary sacrifice represent graded causal pressure, not automatic activation. In the revised simulation, C6 is activated by a sigmoidal function only when C1 is present as an enabling condition and the accumulated input exceeds the activation threshold. Therefore, C1 is necessary but not sufficient; threat, anticipated collective harm, and the moral goal of minimizing deaths must jointly contribute to the activation of sacrifice.

2.1.6. Simulated Scenarios

The system's baseline dynamics were initially defined by the equation $V(t + 1) = f[V(t) \times W]$, where V represents the vector of node activations at time t and W is the influence matrix. To address the loss of intensity produced by the binary function, the revised model specifies a comparative sigmoidal activation criterion: $g(x) = 1/(1 + e^{\{-\lambda(x - \theta)\}})$, where λ controls the slope and θ represents the activation threshold. For interpretability, node activations above the threshold are treated as substantively active, while intermediate activations are interpreted as partial causal pressure. In addition, C6 was modeled with a necessary-condition rule: $C6(t + 1) = g(\text{net}C6)$ only if C1 is active; if C1 is absent, $C6(t + 1) = 0$ regardless of the pressure generated by C2, C3, C4, or C5. This correction prevents war pressure alone from producing voluntary sacrifice and prevents C1 alone from being interpreted as sufficient for the complete sacrificial configuration.

Four scenarios were simulated: 1) complete conflict, with C1, C2, and C3 active at the start; 2) internal causal condition only, with C1 active; 3) war without internal condition, with C2 and C3 active; and 4) empty system, with no initially active nodes. Each scenario was iterated until stability, absence of relevant activation, or convergence toward a stable activation pattern was observed. The revised stability criterion considered both repeated activation tendencies and whether the activation of C6 met the necessary-condition and threshold requirements.

3. Results

3.1. Result 1: Cognitive Architecture of the Decision

The cognitive task analysis identified that the behavior depends on a sequence of

operations transforming a threat state into a sacrifice strategy. The *first operation* is recognizing the conflict configuration: active antagonist, ongoing war, and potential harm to third parties. The *second operation* is identifying the internal causal condition: the agent recognizes that their own existence is part of the structure maintaining the enemy active. The *third operation* consists of inferring consequences: if no action is taken, the war continues; if the internal condition is eliminated, the enemy weakens; if the enemy weakens, the probability of future deaths decreases. The *fourth operation* involves applying a moral criterion: minimizing others' deaths becomes a priority over self-preservation. The *fifth operation* converts the strategy into action: voluntary surrender as a causal intervention.

This result shows that the decision is not explained by a single causal node. The internal condition is necessary to open the possibility of sacrifice as a strategy but is not sufficient by itself to justify the behavior. Active threat, conflict continuity, anticipation of future deaths, and the moral goal of harm minimization operate jointly. In the absence of this integration, sacrifice could appear irrational or narratively arbitrary; within the model, it becomes a coherent strategy under explicit assumptions.

3.2. Result 2: Affective-Moral Structure

The affective-moral appraisal structure showed a hierarchy in which collective goals dominated over individual survival. The condition of carrying an internal causal element produced a negative appraisal regarding self-preservation but simultaneously activated a strategic opportunity. Ongoing war and future deaths generated concern and empathy, emotions associated with the urgency to act. Sacrifice produced an ambivalent appraisal: negative due to its individual cost, positive due to its causal utility. The possibility of weakening the enemy and ending the conflict generated hope and anticipated relief.

This affective configuration was not interpreted as a set of independent emotional labels but as a valuation network that increases the strength of morally regulated action. Concern and empathy increase the weight of goal C5; hope increases the plausibility of C6; and anticipated relief reinforces C7 as a desirable outcome. Together, emotion functions as a meaning organizer and not as a substitute for reasoning.

3.3. Result 3: Dynamics of the Fuzzy Cognitive Map

In the complete conflict scenario, the initial vector activated C1, C2, and C3. Under the revised thresholded-sigmoidal interpretation, these inputs jointly increased the activation pressure on future deaths, the goal of minimizing deaths, voluntary sacrifice, and weakening the antagonist. Unlike the binary version, the corrected simulation treats C6 as substantively active only when the internal causal condition is present and the accumulated pressure exceeds the activation threshold. Therefore, the stable interpretive core is C5-C6-C7, but it emerges from the interaction among C1, threat, conflict continuity, expected harm, and moral

self-regulation rather than from a single causal node.

In the internal-condition-only scenario, C1 generated partial causal pressure toward weakening the antagonist and made sacrifice conceptually possible, but it did not by itself constitute a sufficient condition for full activation of C6. This corrected interpretation directly addresses the over-weighting observed in the binary version: carrying the internal causal condition is necessary for sacrifice to be meaningful, but the action becomes coherent only when the condition is embedded in a broader configuration of active threat, future harm, and the moral priority of minimizing deaths.

In the war without internal condition scenario, threat and conflict activated future deaths and the goal of minimizing deaths, but the absence of C1 blocked the activation of voluntary sacrifice as a causally justified strategy. Thus, C2, C3, and C4 may produce moral pressure and urgency, but they do not generate C6 when the necessary internal causal condition is missing. Finally, in the empty system, no relevant activation was observed, confirming that the model does not produce behavior without causal input. **Table 4** shows the corrected interaction of these scenarios.

Table 4. Corrected scenarios and interactions.

Scenario	Initial Vector	Observed Stable State	Interpretation
Complete conflict	C1 = 1, C2 = 1, C3 = 1	C5, C6, and C7 form the stable interpretive core, with C1 operating as a necessary enabling condition.	Threat, internal causal condition, and conflict jointly converge toward harm minimization, sacrifice, and weakening of the antagonist.
Internal causal condition only	C1 = 1	C1 produces partial causal pressure; C6 remains below the substantive activation threshold.	The internal condition is necessary but not sufficient; the result no longer supports a full sacrificial trajectory in the absence of threat and expected collective harm.
War without internal condition	C2 = 1, C3 = 1	C5 may activate as moral pressure; C6 = 0 because C1 is absent.	War pressure and anticipated deaths do not justify sacrifice unless the internal causal condition is present.
Empty system	All nodes = 0	No activation	The model does not produce behavior without causal input.

3.4. Result 4: System Equilibrium Core

The most consistent interpretive pattern was the core C5-C6-C7, provided that C1 is present as an enabling condition and that threat/conflict inputs supply sufficient activation pressure. This core represents the convergence among a moral goal, a high individual cost action, and a strategic outcome. C5 maintains active orientation to reduce harm; C6 operationalizes the action that modifies the system; and C7 represents the outcome that reduces the threat. In the corrected model, the core's stability suggests that the final behavior emerges when the inter-

nal causal condition, the moral goal, the action, and the expected weakening of the antagonist mutually constrain one another. From a cognitive engineering perspective, this convergence is relevant because it transforms a sacrifice narrative into an explicit causal architecture.

4. Discussion

The objective of this article was to build a model integrating cognitive analysis, affective-moral appraisal, and causal dynamics to explain a voluntary sacrificial decision. The results support the working hypothesis: the action does not behave as a linear product of a single cause but as an emergent property of a system of relationships. The convergence toward C5-C6-C7 shows that the final behavior depends on the interaction among moral goal, strategic action, and expected outcome. This is consistent with the idea that moral judgment combines affect, reasoning, and consequence representation [2] [3]. The revised thresholded interpretation strengthens this claim by preventing voluntary sacrifice from being generated either by C1 alone or by war pressure alone.

The main contribution of the study is methodological. The model offers a procedure to translate narrative behavior into a formal system. First, it identifies conditions and cognitive operations; second, it organizes goals and appraisals; third, it constructs a causal matrix; and fourth, it observes the system's evolution. This sequence is especially useful for problems where behavior is too complex to be explained by simple linear relationships. Instead of asking only for the cause of the decision, the model asks for the minimal configuration that allows an extreme action to be cognitively coherent. This contribution is specifically methodological: the study provides a bridge from qualitative cognitive-affective analysis to a formal dynamic representation whose assumptions can be inspected, revised, and simulated.

The analysis also contributes results to research on sacrificial dilemmas. The modeled behavior superficially approximates sacrificial-type dilemmas but is not equivalent to them. In many experimental dilemmas, it is evaluated whether it is permissible to harm one person to save several. In the analyzed case, the agent decides to assume harm upon themselves. This difference is crucial because it modifies the attribution of agency, responsibility, and moral legitimacy. Kahane *et al.* [1] warn that sacrificial dilemmas should not be automatically equated with general utilitarianism. In this study, the decision is interpreted as self-referential sacrifice regulated by causal responsibility, not as instrumental harm imposed on a third party.

The affective-moral structure reinforces this interpretation. Emotions do not appear as irrational triggers but as indicators of relevance and urgency. Concern for conflict continuity, empathy for future deaths, and hope associated with the possibility of resolution fulfill organizing functions. This reading coincides with Scherer and Moors's perspective [10], for whom emotional processes are linked to evaluation of goals, implications, and action possibilities. Consequently, sacri-

fice does not arise from emotion suppression but from integration among emotion, goal, and causal reasoning. In the FCM, these appraisals do not function as separate emotional labels; they are operationalized as causal pressures that increase or constrain specific nodes and edges, especially $C4 \rightarrow C5$, $C5 \rightarrow C6$, $C6 \rightarrow C7$, and $C7 \rightarrow C5$.

Fuzzy cognitive maps proved suitable to represent this integration because they allow modeling feedback holistically. Action C6 was not treated as an endpoint but as a node that transforms the system. This treatment is theoretically richer than a linear sequence because it recognizes that the agent can anticipate the effects of their own action on the environment. Recent literature on fuzzy cognitive maps has highlighted precisely this capacity to represent expert knowledge and dynamic scenarios in complex contexts [11] [12]. The revised activation criterion also clarifies that causal influence and substantive activation are not equivalent: a node may receive positive input without reaching the threshold required to count as a stable behavioral component.

The C1-only scenario must therefore be interpreted carefully. The current corrective output separates the model's computational behavior from the intended theoretical condition set: C1 defines the causal relevance of sacrifice, but active threat, expected harm, action opportunity, and moral self-regulation define the conditions under which sacrifice becomes behaviorally coherent. This distinction preserves the central claim that sacrificial behavior is an emergent configuration, not a direct consequence of one isolated internal condition.

5. Conclusions

The study developed a cognitive-affective and dynamic model to analyze a voluntary sacrificial decision in an extreme moral context. The integration of cognitive task analysis, affective-moral appraisal, and fuzzy cognitive maps allowed representing the behavior as the result of interaction among environmental conditions, causal inferences, collective goals, evaluative emotions, and strategic action.

The results show that, in the complete conflict scenario and under the corrected thresholded interpretation, the system converges toward a core composed of minimizing deaths, voluntary sacrifice, and weakening the antagonist. This convergence suggests that the final action can be interpreted as an emergent property of a causal network and not as an impulsive response nor as an isolated choice. The revised model also differentiates necessary and sufficient conditions more explicitly: C1 is necessary for the sacrificial strategy to be causally meaningful, but it is not sufficient without threat, expected collective harm, and moral self-regulation.

However, the results also show internal limitations. Although the revised sigmoidal and thresholded criterion reduces the overactivation of C1 and preserves activation intensity, the specific numerical parameters and causal weights remain modeling assumptions. A second limitation is the absence of independent expert validation of the matrix; in the present version, the weights are documented as author-consensus analytical coding derived from the cognitive task analysis and

appraisal phases. A third limitation is the absence of empirical data. The study does not measure real decisions, reported emotions, or behavior under pressure. Therefore, its conclusions should be understood as modeling propositions. This point is fundamental to meet specialized research standards.

Despite these limitations, the model has applied implications. In education, it can be used to teach analysis of complex decisions and distinguish among facts, inferences, values, emotions, and actions. In artificial intelligence, it can serve to discuss how to represent moral reasoning without reducing it to rigid rules. In educational psychology and professional training, it can support the design of scenarios where students identify breaking points, inference errors, and value conflicts. In cognitive engineering, it offers an example of how to convert qualitative knowledge into a simulable architecture. Future research is recommended to validate the matrix through expert judgment, estimate inter-rater agreement for nodes and causal weights, compare activation functions, and apply the model to other moral decision cases, both narrative and professional.

Acknowledgements

This work represents the research conducted by Cecilia Jokebed Rea Castañares to obtain a Master's degree in Optimization from Universidad Autónoma Metropolitana-Azcapotzalco. It is also part of the divisional project, "Design of Intelligent Interfaces for Simulating the Behavior of Living or Animate Organisms," from the same University.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kahane, G., Everett, J.A.C., Earp, B.D., Caviola, L., Faber, N.S., Crockett, M.J., *et al.* (2018) Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology. *Psychological Review*, **125**, 131-164. <https://doi.org/10.1037/rev0000093>
- [2] Greene, J.D. (2015) The Rise of Moral Cognition. *Cognition*, **135**, 39-42. <https://doi.org/10.1016/j.cognition.2014.11.018>
- [3] Crockett, M.J. (2016) How Formal Models Can Illuminate Mechanisms of Moral Judgment and Decision Making. *Current Directions in Psychological Science*, **25**, 85-90. <https://doi.org/10.1177/0963721415624012>
- [4] Bostyn, D.H., Sevenhant, S. and Roets, A. (2018) Of Mice, Men, and Trolleys: Hypothetical Judgment versus Real-Life Behavior in Trolley-Style Moral Dilemmas. *Psychological Science*, **29**, 1084-1093. <https://doi.org/10.1177/0956797617752640>
- [5] Francis, K.B., Howard, C., Howard, I.S., Gummerum, M., Ganis, G., Anderson, G., *et al.* (2016) Virtual Morality: Transitioning from Moral Judgment to Moral Action? *PLOS ONE*, **11**, e0164374. <https://doi.org/10.1371/journal.pone.0164374>
- [6] Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., *et al.* (2018) The Moral Machine Experiment. *Nature*, **563**, 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- [7] Bonnefon, J., Shariff, A. and Rahwan, I. (2016) The Social Dilemma of Autonomous

- Vehicles. *Science*, **352**, 1573-1576. <https://doi.org/10.1126/science.aaf2654>
- [8] Ortony, A., Clore, G.L. and Collins, A. (1988) The Cognitive Structure of Emotions. Cambridge University Press. <https://doi.org/10.1017/cbo9780511571299>
 - [9] Ortony, A., Clore, G.L. and Collins, A. (2022) The Cognitive Structure of Emotions. 2nd Edition, Cambridge University Press. <https://doi.org/10.1017/9781108934053>
 - [10] Scherer, K.R. and Moors, A. (2019) The Emotion Process: Event Appraisal and Component Differentiation. *Annual Review of Psychology*, **70**, 719-745. <https://doi.org/10.1146/annurev-psych-122216-011854>
 - [11] Amirkhani, A., Papageorgiou, E.I., Mohseni, A. and Mosavi, M.R. (2017) A Review of Fuzzy Cognitive Maps in Medicine: Taxonomy, Methods, and Applications. *Computer Methods and Programs in Biomedicine*, **142**, 129-145. <https://doi.org/10.1016/j.cmpb.2017.02.021>
 - [12] Felix, G., Nápoles, G., Falcon, R., Froelich, W., Vanhoof, K. and Bello, R. (2019) A Review on Methods and Software for Fuzzy Cognitive Maps. *Artificial Intelligence Review*, **52**, 1707-1737. <https://doi.org/10.1007/s10462-017-9575-1>
 - [13] Aminpour, P., Gray, S.A., Jetter, A.J., Introne, J.E., Singer, A. and Arlinghaus, R. (2020) Wisdom of Stakeholder Crowds in Complex Social-Ecological Systems. *Nature Sustainability*, **3**, 191-199. <https://doi.org/10.1038/s41893-019-0467-z>
 - [14] Gray, S.A., Gray, S., De Kok, J.L., Helfgott, A.E.R., O'Dwyer, B., Jordan, R., *et al.* (2015) Using Fuzzy Cognitive Mapping as a Participatory Approach to Analyze Change, Preferred States, and Perceived Resilience of Social-Ecological Systems. *Ecology and Society*, **20**, Article No. 11. <https://doi.org/10.5751/es-07396-200211>
 - [15] Kosko, B. (1986) Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies*, **24**, 65-75. [https://doi.org/10.1016/s0020-7373\(86\)80040-2](https://doi.org/10.1016/s0020-7373(86)80040-2)
 - [16] Glykas, M. (2010) Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications. Springer.
 - [17] Papageorgiou, E.I. (2012) Learning Algorithms for Fuzzy Cognitive Maps—A Review Study. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **42**, 150-163. <https://doi.org/10.1109/tsmcc.2011.2138694>
 - [18] Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M. and Cohen, J.D. (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, **293**, 2105-2108. <https://doi.org/10.1126/science.1062872>
 - [19] Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. and Cohen, J.D. (2004) The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron*, **44**, 389-400. <https://doi.org/10.1016/j.neuron.2004.09.027>
 - [20] Haidt, J. (2001) The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, **108**, 814-834. <https://doi.org/10.1037/0033-295x.108.4.814>
 - [21] Laureano-Cruces, A.L. (2024) Methodology for the Analysis and Design of a Cognitive Model: Emotions and Artificial Conscience Simulation. BP International.
 - [22] Moll, J., de Oliveira-Souza, R., Bramati, I.E. and Grafman, J. (2002) Functional Networks in Emotional Moral and Nonmoral Social Judgments. *NeuroImage*, **16**, 696-703. <https://doi.org/10.1006/nimg.2002.1118>
 - [23] Rowling, J.K. (2007) Harry Potter and the Deathly Hallows. Bloomsbury.