2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS) | 978-1-6654-5238-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICHMS56717.2022.9980746

Fuzzy Mental Model Finite State Machines: A Mental Modeling Formalism for Assessing Mode Confusion and Human-machine "Trust"

Matthew L. Bolton
Engineering Systems and Environment
University of Virginia
Charlottesville, USA
mlb4b@virginia.edu

Elliot Biltekoff
Industrial and Systems Engineering
University at Buffalo
Buffalo, USA
elbiltek@buffalo.edu

Kevin Byrne
Engineering Systems and Environment
University of Virginia
Charlottesville, USA
kcx2zj@virginia.edu

***Abstract*—Formal human-machine analyses based around human mental models have shown great utility for discovering when and how people may develop mode confusion and be surprised or disoriented by automated behavior. Such analyses represent mental models with finite s tate m achine formalisms. These are limited in that they assume unrealistic precision in how people think about states and input events. This paper proposing a new formalism called Fuzzy Mental Model Finite State Machines (FMMFSMs). FMMFSMs combine state machine modeling with fuzzy logic to allow for precise reasoning about the imprecision of human mental model states and inputs. This has the potential to enable formal mental model analyses to support traditional mode confusion, but also account for things like drift: where the humans mental model changes over time due to stagnant or slowly changing conditions. This paper presents the FMMFSMs formalism and illustrates is potential for finding m ode confusion, automation surprise, and trust disruption with an automobile automation application. Implications of these developments and future research are discussed.**

***Index Terms*—Formal methods, mental model, mode confusion, automation surprise, fuzzy logic, state machine, trust**

I. Introduction

A mental model is "a user's internal representation of the function of the target system" [1, p. 7]. Critically, when a person is interacting with a system, they will "run" a mental model to track the system state and anticipate future behaviors. Mental models can be abstract, incomplete, unscientific, and imprecise [1]. Enabling people to maintain accurate mental models is critical to safe and effective human-machine interaction. A lack of correspondence between a human's mental model and the actual system constitutes mode confusion [2]. Mode confusion can cause automation surprise, where the system does something unexpected or does not do something expected [3]. This can disorient an operator (causing errors) and erode human-automation trust. As such, mental modeling has become instrumental in contemporary design practices for decision support, joint cognitive systems, and human-machine interaction [4]–[7]. Mental models have also been employed in formal verification. In this, automated proof techniques mathematically find system and mental model inconsistencies, automation surprise, human errors, and engineering interventions to address them [8]–[10].

Formal analyses typically represent mental models as non-deterministic state machines where events (human actions or environmental changes) cause state transitions. These models imply binary/crisp human perception of transition events (they either happened or did not) and current state (the model is in a given state, or it is not). Thus, this formalism can support abstract, incomplete, and unscientific mental models, but not imprecise ones. As such, traditional state machines provide no insights into the degree of automation surprise and associated trust degradation.

II. Objective

We propose addressing these shortcomings with Fuzzy Mental Model Finite State Machines (FMMFSMs). This new formalism uses state machines as its base but allows for possibilistic uncertainty using fuzzy logic [11]. Fuzzy sets and logic were developed to naturally capture the vagueness inherent to human thinking: things do not necessarily fall into discrete categories (e.g., true or false) but rather into all categories with degrees of membership (fuzzy set memberships with degrees from 0, not at all, to 1, completely).

Note that fuzzy set membership differs from probability. Probability is associated with providing precise measures of the likelihood of an event occurring. Conversely, fuzzy set membership precisely captures how vague something is. This is why fuzzy set membership is a better candidate for capturing the imprecision/vagueness of human mental models. This distinction is important because fuzzy set memberships and probabilities have similar, but different, mathematical properties. For example, the fuzzy set memberships around related labeled concepts (e.g. true or false) need not sum to one.

We hypothesize that, through FMMFSMs, fuzzy logic can be used with state machines to precisely represent mental model imprecision. In this paper, we present a definition of the FMMFSM formalism. We then use a simple driving automation case study to illustrate its potential utility for predicting human mode confusion and the degree of automation surprise.

III. Fuzzy Mental Model Finite State Machines (FMMFSMs)

FMMFSMs have been conceptualized to capture uncertainty that could exist in both human event perception and the state the human thinks the system is in. The fuzziness of both is

978-1-6654-5238-0/22/$31.00 ©2022 IEEE

accounted for in state transitions. Thus, at any operational instance, a FMMFSM should describe how possible the person thinks it is that the system is in any state. The model will logically transition to a new distribution of fuzzy state memberships based on the current distribution and the fuzziness of perceived events. This formalism should thus allow analysts to quantify the degree (if any) of the discrepancy between (fuzzy) mental and (crisp) system model states based on the distance between memberships of the actual system state (0 for not in the state or 1 for being in the state) and its corresponding mental model state (a value inclusively between 0 and 1).

In what follows, let $\mathbb{F}$ be the set of possible fuzzy membership values: $\mathbb{F} = \{x : x \in \mathbb{R} \wedge x \in [0,1]\}$.

A FMMFSM is an 8-tuple: $(\Sigma, Q, I, M, \vec{m}_0, \alpha, \phi, \delta)$:

- Σ is a finite set of input events (human actions, environmental events, etc.) that can cause a change in system state that are observable by a person.
- Q is a finite set of states the person thinks the system could be in.
- I is the set of all possible vectors of fuzzy membership values of input events from Σ. That is, each entry in a given vector $\vec{i} \in I$ represents the degree to which the person thinks that the associated input event occured: $\forall \vec{i} \in I, (|\vec{i}| = |\Sigma|) \wedge (\forall i \in \vec{i}, i \in \mathbb{F})$. Note that vertical bars $||$ represent the cardinality of the contained vector or set.
- M is a set of all possible vectors of fuzzy membership values for each state from Q. That is, each entry in a given vector $\vec{m} \in M$ represents the degree to which the person thinks the machine is in the associated state of Q: $\forall \vec{m} \in M, (|\vec{m}| = |Q|) \wedge (\forall m \in \vec{m}, m \in \mathbb{F})$.
- $\vec{m}_0 \in M$ is a vector representing initial state fuzzy set memberships: the degree to which the person thinks the machine is initially in each state.
- $\alpha : \Sigma \rightarrow I$ is a function that describes how inputs are fuzzified by mapping crisp input events from Σ to a vector of input fuzzy set memberships from I.
- $\phi : Q \times \Sigma \rightarrow M$ is a function that describes the fuzzification of state transitions (the degreee to which the person thinks that a given state and input will transition to another given state) by mapping current states (Q) and input events (Σ) to a vector of membership values for next states (M). Thus, $\forall s \in \Sigma$ and $\forall q \in Q, \phi(q,s)$ describes the degree to which the person thinks the machine will be in each state of Q in the next state following input event s when the machine is in state q.
- $\delta : M \times \Sigma \rightarrow M$ is the "next state" function. This uses α and ϕ to describe how a given vector of fuzzy state memberships (the degree to which the person thinks that the machine is in each state) will change in response to an input event. If we let $\vec{m}' \in M$ be the resulting "next state" vector of fuzzy state memberships associated with a given event $s \in \Sigma$ happening with a current state memberships vector $m \in M$ and we let V_i generically represent the value associated with a given item i in any given vector V, then:

$$\delta(\vec{m}, s) = \vec{m}' \text{ where } \forall m'_x \in m' : \quad m'_x = \bigvee_{m_y \in \vec{m}, s_z \in \alpha(s)} (m_y \wedge s_z \wedge \phi(y,z)_x). \quad (1)$$

The rationale for the expression in Eq. (1) follows. For any given state $y \in Q$ and input event $z \in \Sigma$, membership in a next state $x \in Q$ will be present **IF** the state machine is in y **AND** the event z occurs **AND** the state machine can transition to x from y on z. Thus, the fuzzy membership value of x in the next state as contributed by this condition will be the fuzzy **AND** of current state membership (m_y), input event membership (s_z), and membership of x produced from y transitioning on z ($\phi(y,z)_x$). Because any possible combination of current states and inputs could transition to x, we take a fuzzy **OR** of all possible memberships of current state and input event combinations transitioning to x.

There are different ways for computing **AND** ($\wedge$) and **OR** ($\vee$) on fuzzy set memberships [12]. This includes the standard

$$\bigwedge_{i=1}^{n} x_i = \text{Min}(x_1, x_2, ..., x_n) \quad (2)$$

and

$$\bigvee_{i=1}^{n} x_i = \text{Max}(x_1, x_2, ..., x_n). \quad (3)$$

However, we are interested in how fuzzy state memberships could evolve over time (e.g., how the membership degree of a given mental model state will reduce the longer contradictory evidence of that state is perceived). This is not possible with Eqs. (2) and (3) because they only select from the set of considered fuzzy numbers (i.e., the exact values of x_1, x_2, ..., x_n). Thus, we used the alternative formulations [12] of

$$\bigwedge_{i=1}^{n} x_i = \prod_{i=1}^{n} x_i \quad (4)$$

and

$$\bigvee i = 1^n x_i = 1 - \prod_{i=1}^{n} (1 - x_i), \quad (5)$$

which are more similar to logical operations on probabilities.

IV. Illustrative Example

To illustrate the potential utility of this formalism for finding mode confusion and automation surprise, we now apply it to a very simple automobile automation application (inspired by cases from [13], [14]), a cruise control.

For the sake of simplicity, we assume that this system has two states: Active, the cruise control is enabled at a given speed, and Inactive, the cruise control is not initiated and will not control the speed of the car. We also simplistically assume that the system can respond to two inputs: the human applying the Gas and the human not applying it (NoGas). The way the automobile drives will depend on the unique combinations of the system state and inputs. When the cruise control is Inactive, the car will be propelled forward at a speed and acceleration commensurate with the application of Gas. If there is NoGas, the car will not be driven at all by the engine and decelerate in most situations. When the cruise control is Active, the automobile will maintain its current speed when NoGas is applied. However, the driver has the option to override the cruise control by pressing the Gas to achieve speeds beyond what was set. In this situation, if the driver ceases to apply gas

(NoGas occurs), the vehicle should (in most conditions) slow down to the speed set in the cruise control. However, once the car hits the cruise speed, it should proceed at that speed with no additional deceleration.

To model this application as a FMMFSM, we first define sigma based on the human action input events that can occur:

$$\Sigma = \{\text{Gas}, \text{NoGas}\}. \tag{6}$$

Second, we can define the system states:

$$Q = \{\text{Active}, \text{Inactive}\}. \tag{7}$$

I and M are thus defined by the possible vectors of membership values for each element in Σ and Q respectively:

$$I = \{(\ \overset{\text{Gas}}{i_{\text{Gas}}},\ \overset{\text{NoGas}}{i_{\text{NoGas}}}\ ) | i_{\text{Gas}} \in \mathbb{F} \wedge i_{\text{NoGas}} \in \mathbb{F}\} \tag{8}$$

and

$$M = \{(\ \overset{\text{Active}}{m_{\text{Active}}},\ \overset{\text{Inactive}}{m_{\text{Inactive}}}\ ) | m_{\text{Active}} \in \mathbb{F} \wedge m_{\text{Inactive}} \in \mathbb{F}\}. \tag{9}$$

Our analysis starts assuming the person has been driving with cruise activated with NoGas. Thus, the driver is relatively certain that cruise is Active:

$$\vec{m}_0 = (\ \overset{\text{Active}}{1},\ \overset{\text{Inactive}}{0.01}\ ). \tag{10}$$

Additionally, we assumed that the driver would be relatively certain each of the two actions when being applied, where there may be slight uncertainty about Gas application (e.g., a foot is against the pedal, but not applying effective pressure). Thus, we assigned an α function as shown in Table I. We also used our intuition to assign fuzzy membership values to the conditions required for ϕ (Table II). Note that in these cases, the driver will be most certain that the automobile will remain in its current state regardless of the action. There is more uncertainty when gas is being applied because the car dynamics will not indicate whether the cruise control is active. Finally, δ is defined as specified by Eq. (1) using the respective definitions of fuzzy $\wedge$ and $\vee$ operations from Eqs. (2) and (3).

To illustrate how this model could be used to find mode confusion, we can simulate how the human's concept of state membership changes in response to events and over time. For this, we assumed that at the beginning of our analysis, the driver (who had been previously driving with NoGas and cruise Active; Eq. (10)) applies Gas to respond to conditions in the environment (such as speeding up to accommodate a merging car). We assume that the driver then continues applying Gas

TABLE I
α FOR THE CRUISE CONTROL FMMFSM

Event	Vector (Gas, NoGas)
Gas	(0.99, 0.01)
NoGas	(0.00, 1.00)

TABLE II
ϕ FOR THE CRUISE CONTROL FMMFSM

State	Event	Vector (Active, Inactive)
Active	Gas	(0.9, 0.2)
Inactive	Gas	(0.2, 0.9)
Active	NoGas	(0.9, 0.1)
Inactive	NoGas	(0.0, 1.0)

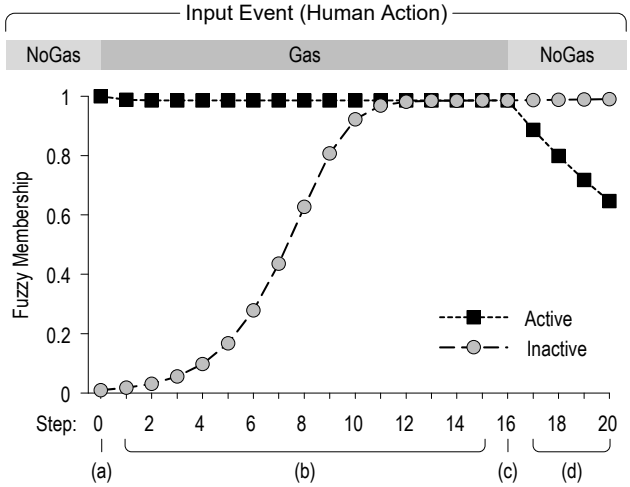

Fig. 1. Graph showing how human fuzzy membership values for cruise control Active and Inactive states (as predicted by the FMMFSM described in Eqs. (6) to (9) and Tables I and II) change over model steps through a scenario. (a) The cruise control starts active with the human having a fairly accurate assessment of cruise state just as the gas is applied. (b) The human's understanding of the possible cruise states becomes more ambiguous the longer the human applies the gas. (c) Inactive membership exceeds Active membership just as the human releases the gas. (d) The human feels the car decelerate and Active membership precipitously drops. This constitutes a potentially dangerous mode confusion, where the human inaccurately and increasingly thinks that it is more Possible that cruise is Inactive than Active.

(controlling the car's speed above the one set for cruise) for 16 model steps. Figure 1 shows how the Active and Inactive state membership values change/converge over this period, with Active membership slowly decreasing and Inactive membership increasing until it overtakes Active at step 16.

While this constitutes a potential mode confusion condition, where the driver effectively thinks that it is slightly more possible that the cruise control is inactive than active, the situation can get worse. From here, we assumed that the driver removed their foot from the gas (applied NoGas; e.g. to slow while driving up an exit ramp). As Fig. 1 shows in steps 16–20, this causes the membership of Active to quickly drop while membership in Inactive continues to slowly rise. This could produce a potentially dangerous automation surprise for a driver who may find their car suddenly applying gas in a situation where they need to slow down, such as coming into a tight exit ramp curve or approaching a stop.

V. DISCUSSION

In this work, we have introduced the new FMMFSMs formalism for modeling human mental models. This combines concepts from fuzzy logic and state machines. This enables a precise representation of the vagueness humans will have about system input events and the state a given automated system is in. The presented application demonstrates how this formalism could be useful for reasoning about mode confusion and automation surprise. In particular, it provides an unprecedented ability to examine the degree to which human interpretation of system sate can diverge from reality, and how this divergence can drift or aggregate over time.

The magnitude of divergence provides a first step towards quantifying the degree of mode confusion. This could provide

insights into the level of surprise and disruption such mode confusion could manifest. It could also potentially allow for a quantitative understanding of how automation surprise could damage human-machine trust. However, it is important to mind that the research presented here is preliminary. As such, there is much potential for future research.

We fully acknowledge that the application we evaluated is simplistic and potentially unrealistic. This is because it was intended to be illustrative rather than representative. However, there are real examples of the types of automation surprises exhibited by cruise controls [13]. Thus, our example is not completely artificial. In any case, future work will investigate more complete, realistic, and validated applications such as the complete cruise control systems from Lee et al. [15] or more "cutting-edge" automobile automation.

One of the biggest challenges of modeling with fuzzy logic is eliciting the fuzzification processes. This is because they traditionally require direct collection from humans [16]. Similar constraints impact the identification of state-based human mental models. Future work should investigate how to efficiently collect such information from humans and/or develop algorithms that could learn it from data.

It is worth noting that the fuzzification process presented in the cruise control application are simplistic. In more standard fuzzy analyses, a membership function determines how crisp, measurable quantities from the environment are fuzzified. Membership function implementation should be compatible with the formulation of α from Section III. This should be investigated in more depth in future research.

Fuzzy state membership values of FMMFSMs have potential for providing model-based predictions about the degree of human mode confusion, automation surprise, and human-machine trust. However, this connection will need to be validated. Future research should empirically investigate how fuzzy state membership values correspond with these phenomena.

Finally, to be genuinely useful for finding unexpected human-machine interaction problems, FMMFSMs should be adapted for use in model-based analyses. Our case study shows that FMMFSMs are compatible with simulation. More complete evaluations could be achieved by adapting FMMFSMs for use in formal analyses like model checking [17]: an automated approach to formal verification. Model checking has been used successfully to explore mode confusion and automation surprise (see [18]) because proves properties about state-machines. Operations on fuzzy memberships are incompatible with standard model checkers because they are continuous and require nonlinear operations (e.g. Eqs. (4) and (5)). However, fuzzy memberships and their operations are very similar to those of probabilities, and there are model checkers that can handle stochastics [19]. Future research should investigate how to adapt these tools for use with FMMFSMs.

REFERENCES

[1] D. A. Norman, *Some observations on mental models*. Erlbaum: Hillsdale, NJ, 1983, ch. 1, pp. 7–14.

[2] N. B. Sarter and D. D. Woods, "How in the world did we ever get into that mode? Mode error and awareness in supervisory control," *Human Factors*, vol. 37, no. 1, pp. 5–19, 1995.

[3] E. Palmer, "'Oops, it didn't arm'- A case study of two automation surprises," in *Proceedings of the 8th International Symposium on Aviation Psychology*. Dayton: Wright State University, 1995, pp. 227–232.

[4] A. Houser, "Mental models for cybersecurity: A formal methods approach," Ph.D. dissertation, University at Buffalo, the State University of New York, Buffalo, 2018.

[5] E. Hollnagel and D. D. Woods, *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press, 2005.

[6] D. D. Woods and E. Hollnagel, *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press, 2006.

[7] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering Psychology & Human Performance*. Psychology Press, 2015.

[8] A. Degani and M. Heymann, "Formal verification of human-automation interaction," *Human Factors*, vol. 44, no. 1, pp. 28–43, 2002.

[9] J. Bredereke and A. Lankenau, "A rigorous view of mode confusion," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2002, pp. 19–31.

[10] S. Combéfis and C. Pecheur, "Automatic generation of full-control system abstraction for human-machine interaction," *Human-Machine Interaction (Formal H)*, p. 9, 2012.

[11] L. A. Zadeh, "Fuzzy sets," in *Fuzzy sets, fuzzy logic, and fuzzy systems: Selected papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.

[12] M. J. Wierman, "An introduction to the mathematics of uncertainty," *Creighton University*, pp. 149–150, 2010.

[13] A. Degani, *Taming HAL: Designing interfaces beyond 2001*. Springer, 2004.

[14] M. L. Bolton, R. I. Siminiceanu, and E. J. Bass, "A systematic approach to model checking human–automation interaction using task analytic models," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 5, pp. 961–976, 2011.

[15] S. H. Lee, D. R. Ahn, and J. H. Yang, "Mode confusion in driver interfaces for adaptive cruise control systems," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014, pp. 4105–4106.

[16] L. A. Zadeh, *Computing with words: Principal concepts and ideas*. Springer, 2012, vol. 277.

[17] E. Clarke, O. Grumberg, and D. Peled, *Model Checking*. MIT Press, 1999.

[18] M. L. Bolton, E. J. Bass, and R. I. Siminiceanu, "Using formal verification to evaluate human-automation interaction: A review," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 3, pp. 488–503, 2013.

[19] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *Proceedings of the 23rd International Conference on Computer Aided Verification*, ser. Lecture Notes in Computer Science, vol. 6806. Springer, 2011, pp. 585–591.