OSC Model Introduction

We begin our discussion of the models we've selected to compose our composite concept model with the 'Ochs Sabouret Corruble' Model, which we named after its principal authors (and shortened to the OSC model for short as will be used hereafter). The OSC Model encompasses a "model of personality, emotions and social relations aiming at the improvement of NPC credibility in video games", and the concept of credibility is crucial both WRT this model as well as for NPC AI in general. Credibility speaks for how immersive a video game or otherwise simulation is in support of what the authors call "the willing suspension of disbelief [with respect to] the degree of consistency" between all components which make up the environment (i.e. terrain, buildings, NPCs, etc.) Indeed, we have already explored and discussed a similar idea of consistency in our work with Procedural Content Generation when introducing the 'Principle of Physical Stability and Contextual Consistency' as a qualification thereof regarding the generation of a simulated world (sans NPCs). Thus, in a sense, this report could be seen as a corollary to the work done in Project Genesis; especially as a definition for NPC Credibility and thus a measure of consistency is not only provided but likewise acts as the foundation by which credibility and immersion for NPCs are measured. In fact, OSC's concept of credibility parallels that of Genesis even more so by defining two important dimensions involving 'Behavioral Consistency' as a means to measure the credibility of an implementation of NPC behavior; to the point that it may even satisfy as the measure of Contextual Consistency in our Genesis model within the domain of NPCs.

The first dimension is 'consistency with past behavior', and simulates human personality and the retention of individual experience; especially in terms of how they direct the NPC's behaviors and preferences/attitudes in the present, and how they themselves evolve in the future. Personality is defined within this dimension as "the set of psychological traits...which impacts the perceptions, motivations and actions of a [NPC]", and Individual Experience as "the set of events the [NPC] lived previously and memorized". From these definitions, we see clear connections to several of the major types of agents studied in this course, specifically the Model-Oriented agents of types Utility-Based, Goal-Based, and Memory-Based (i.e. Knowledge-Based). The second dimension is 'consistency with the current environment' and simulates how human behavior can be instantly reactive to changes in the world at the present. This dimension is critical to credibility when it comes to NPC behaviors on events such as a player suddenly acting violent against an NPC, a disaster or some conflict occurring near to an NPC, and even things as subtle as an NPC's reaction to being provided something it knows to be only a rumor (i.e. non-verified fact). Both dimensions further encompass the simulation of social relationships between NPCs and even unto human players: as both involve the current events of exchanges (i.e. via dialogues between agents) and past exchanges and relations between other agents (i.e. having an existing opinion about a certain character); of which together composes the current 'attitude' that an NPC has with another character.

Background Models for Emotion/Personality/Social-Relations

From these two dimensions: a model can be formed from which a realization of Behavioral Consistency can be implemented - which is exactly what is done by the paper. An important way in which this model satisfies consistency is to derive itself from well-established Psychological models, as well as being designed for straightforward use by game developers (always a welcome idea!) We will discuss these background models before discussing how OSC realizes them in an implementation.

Emotion Model

The Emotion Model encompasses the realization of when certain emotions should appear and in what degree or intensity: as to implement the corresponding means by which they are displayed (e.g. via dialogue selection and facial expression / body gesture animation); and which plans, behaviors, and actions result from them. It uses a small mechanic of behavioral mathematics to further define the "triggering conditions of an emotion" i.e. the circumstances by which they should be perceived by the NPC, and to what degree and intensity; thus realizing ideas like "not sad enough to make a frowning face" or "being extremely scared negatively impacts rational decision-making". A first model from human psychology is Cognitive Appraisal Theory in which "emotions are triggered by subjective appraisal of an event" which alongside perception thereof determines type/intensity of emotion(s). How events are perceived and appraised depends on individual NPC via their goals and beliefs; ergo two agents may experience different emotions to the same event. Following this, OSC makes extensive use of the 'Ortony Clore Collins' Model (hereafter OCC Model) which models the processing of emotions via triggering emotions based on evaluating 3 things: consequences of an event on the agent's goals, the actions performed by the agent, and the objects in the environment. OSC adapts OCC as its main realization based on "attitudes of NPCs towards actions, objects and other characters of the environment." And again – while it suffices for them, we will be spicing things up further with CiF, Needs-Based, etc.

Further: OCC proposes 'set of simple and easily implemented rules' to indicate which emotions are triggered and in which situations; which are described in terms of goals and beliefs. Consequently: it utilizes the Dyer Computational Model wherein "causes of emotions [are] represented by agent's beliefs on state of its goals and on the responsible agent. For instance, agent feels joy when it believes one of its goals is achieved, and anger when another agent threatens one of its goals." "Emotion is triggered by agent's belief that probability to complete one of its goals has been modified. For instance, a positive emotion is triggered when an agent believes that the probability to achieve one of its goals has increased." The final major model discussed is the Pleasure-Arousal-Dominance Model (herafter PAD Model) in which NPC emotions are represented through the 3 emotional dimensions: pleasure, arousal, dominance. To model emotional dynamics, 3 types of influence are considered: physical (i.e. fatigue, illness), chemical (i.e. alcohol, medicine), and psychological (i.e. freedom, hostility) s.t. each type has specific effect on emotional dimensions. For instance, fatigue induces a decrease of pleasure and arousal. OSC Discusses this model mostly as existing work for which they critizisze drawbacks such as how designers must define effect of each event of game on emotional dimensions. While true and what we also want to avoid, we will at least be modelling physical influence in a more procedural manner than naïve 'pre-baked' approaches via Needs-Based AI discussed below in Section 5. Lastly: the classic behavioral mathematics mechanic used to express personality-oriented emotional state is realized with an intensity decay function with activation/deactivation thresholds, as will be discussed later.

The Belief-Desire-Intention Model was briefly discussed (because of course it would show up!) by noting that some computational models use them, but there are goof criticisms thereof. In Particular: similar to the PAD Model: this approach requires a lot of pre-baked definitions to be provided for the NPC's goals, KR, and KB; not to mention given the complexity of larger scaled worlds and unpredictability that comes with agents and players more freely able to interact with it and each other. That Said: one of the models we studied for action/behavior planning, Goal-Oriented Action Planning Model (hereafter GOAP), is a good resolution such that we plan to utilize it for a full realization of Fea. But not for personality sim: and instead as a decision-making, goal-planning, task-scheduling, behavior-constructing, and inference-reasoning co-system of the greater composite model.

Personality Model

When introducing their Personality Model, OSC starts with a good critique of naive methods: “Most games use ad-hoc personality models that not based in Social Science and Psychology” with the same issues aforementioned. To warrant further discussion: Category-based models (in which strongly specific behaviors are linked to categories such as “good/bad”; these categories are defined a priori in the game scripts) and Attribute-based approaches (e.g. for RPG’s: numeric value for ‘charisma’ is used to compute probability that NPC transfers confidential information to PC). This lack of theoretical background often leads to unrealistic behaviors in these games. The OSC model (and our composition) is based on personality traits s.t. such traits affect various behaviors in multiple ways; asserting that they adopt a personality trait-based approach via its credibility as established in Social Science and Psychology.

Therefore, as its primary Personality Model: OSC utilizes the OCEAN i.e. ‘Big 5 Personality Traits’ Model (and we note the same for Needs-Based AI vis-à-vis its greater implementation within The Sims). OCEAN is well-known in both cognitive psychology and among the general public (as seen in countless self-help books). It features a “detailed enumeration of behavior categories attached to each personality trait”, of which such traits of such values are used to help define the emotions and attitudes for NPCs; for which OSC suggests supports NPC credibility via defining emotional reactions WRT both triggered emotions and personality. The 5 Traits are as follows such that each is expressed with its binary counterpart, encompassing a ‘Positive/Negative’ relationship, alongside an equivalent analogy via Wikipedia:

- Extroversion/Introversion: (outgoing/energetic vs. solitary/reserved)
- Neuroticism/Emotional-Stability: (sensitive/nervous vs. secure/confident)
- Agreeableness/Disagreeableness: (friendly/compassionate vs. challenging/detached)
- Conscientiousness/Neglect: (efficient/organized vs. easy-going/careless)
- Open-Mindedness: Openness to experience (inventive/curious vs. consistent/cautious)

OSC ends up only realizing 'Neuroticism' and 'Extroversion' as the others depend on external agents' emotions (perception thereof) else otherwise outside factors not covered by the paper. Fortunately: the CiF Model discussed in the next section appears to fill this gap well, thus our method could certainly realize Agreeableness, likely some degree of Conscientiousness, and possibly even Open-Mindedness!

Social Relations Model

OSC Introduces their social relations model by linking PC interactions with NPC actions (and all other combos) as well as connecting emotion and social dynamics: “Interactions…are not only emotional but also intrinsically social. [NPCs] perform roles portrayed by humans e.g. the role of guide [as described in the introduction section via Star Trek Holodeck and NPCs for pedagogy]. During an exchange, different social relations can then appear between characters (player or non-player)." The various ‘ingredients’ discovered with the paper’s research are as follows, and provides a good recap of some of Fëa’s goals:

- A social-psychological model including personality, emotions, and attitudes such that NPC’s behavior is computed based on its personality and attitudes.
- A Linguistic style of dialog between virtual characters is determined according to social variables (social distance and power) and user’s emotions.
- A Social layer that manages communication and biases plan generation and execution in accordance with a social context.
- Social filters to constrain character’s expression of emotion depending on the social context (represented through social power social distance), NPC’s personality, and emotions.

- Use character's social relations with user to identify the appropriate subject to discuss with the user [We'll also include other NPCs, and this is a major discussion point of Dialogue Agents].

We note that CiF supports ALL of these, especially the latter 3 items for which the OSC model provides only a limited amount of support; as it doesn't tread too deeply into interpersonal / group / tribe / society stuff (but again --- CiF DOES)

OSC realizes a set of 'Social Variables' to represent social relations, wherein each variable characterizes a specific "dimension of social relation between any two agents (virtual or human)". It is similar to the CiF Model discussed in the next section in that it represents such values WRT a bidirectional relationship of which relates {x-to-y, y-to-x}. Further: social relations are not necessarily symmetric! (e.g. 'bully and nerd' vis-a-vis dominance, etc.) They disclaim that such variables are not an exact science, before defining 4 main social variables (plus their negations pairs as per personality traits) used to realize social values as from their cognitive psychology research as follows.

- 'Attitude' represents the 'Degree of Liking' that an NPC has for another NPC or PC
- 'Solidarity' represents the 'Degree of Like-Mindedness' i.e. 'social distance' between two characters (e.g. 'of the same faction')
- 'Dominance' represents the power that an NPC has or can exert upon another NPC (a.k.a. 'intimidation factor')
- 'Familiarity' represents how well an NPC knows another NPC/PC based on the type of info and amount shared

Per usual, the paper reminds of existing work in comparison to the enhanced capabilities of their model; specifically, regarding the importance of 'dynamic social state given social context'. To start: it observes that others are "generally static" in terms of focusing mainly on how social context drives NPC behavior; or to put it in Bayesian Logic terms: ('act sad' | 'at a funeral'). We note a second definition from our observations in that for many of the same models: the context itself does not change unless the player activates trigger flags (e.g. via dialogue choices, quest item completions, etc.) of which either change social state/context directly; else unlock action/dialogue options which change state when selected by the player. Standalone, this is not necessarily a bad mechanic: in fact both the CiF and Dialogue Agents models discussed below implement this, as does the composite model we envision! Instead, the paper provides a corollary critique which qualifies its predecessor: stating "few researchers propose a model of the dynamics of social context"; for which we note is especially the case for social exchanges involving physical gestures (i.e. 'emotes' as implemented in Fable II) and dialogues between characters. This basically also means two things such that one is explicit and the other implicit.

The explicit part is that the personalities of characters themselves aren't much, if at all, factored into how they behave in social exchanges. Indeed, as discussed in the 'Personality Models' subsection above: such consideration has only been realized through primitive attribute values (e.g. 'charisma' discussed above in the 'Personality Models' subsection) or as 'hardcoded compulsions' to change state as activated by unlocking a scripted trigger (e.g. completed a quest item, joined an enemy faction, caught stealing sweetrolls in Skingrad, etc.) In any case: these two means suffer from the same issues discussed throughout this report: the 'gears' encompassing the dynamics of world and its agents are mostly to fully turned only by hardcoded, pre-scripted triggers that are manually activated by the player via either quest progression or dialogue choices (themselves mainly hardcoded, pre-scripted constructs), else increasing some attribute primitive value to a certain number (i.e. dialogue choices locked unless 'charisma' > 7) only on infrequent events such as levelling up in experience. We note that TES-Skyrim, one of the games we use as a 'benchmark' of where industry is at the present time, features all of these mechanisms in its social dynamics and dialogue mechanics.

The implicit part is also a corollary, for which the paper states (paraphrased) '[existing models] propose functions to effect the dynamics of a character's attitude and familiarity toward another according to the valence [i.e. 'combining power' / 'composite sum'] of emotions triggered during the interaction … however, they don't take into account the distinct impacts of different types of emotions on different, possibly unique personality types. That is, they cannot emphasize (e.g. weigh) certain emotions or attitudes over others: which means they cannot implement individual personalities for the same reason that networks (Bayesian, Constraint, Neural, etc.) without weights cannot implement nodes that are intended to be more sensitive than others. For example: while 'shame' and 'insult' may each involve assigning a score of [-20] to the current conversation and affect the target interlocutor's disposition and dialogue tree accordingly: both may also have vastly different impacts on the greater picture of personal and social state for everyone involved in the conversation. 'Shame' might involve a small decrement to the shamed interlocutor's [Neuroticism] level, while effecting a similar small decrement to one or a few disposition scores by the other interlocutors encompassing "I feel a bit more embarrassed against everyone who was there to see me be shamed, and maybe some unhappiness with the person who shamed me". 'Insult' might involve a larger decrement to the shamed interlocutor's [Agreeableness] level, while effecting as severe a decrement against the [Degree Of Liking] scores of the one who insulted them yet perhaps also a similar reduction in [Neuroticism] against the other characters present to hear the insult. And how much each of these values are incremented/decremented, as well as how such change affects their general mood – should [must] be variable for each unique personality configuration.

OSC Architecture Overview

With the Background Models identified and introduced: more specific implementation details can be discussed regarding OSC. We discuss the basic Knowledge Representation (hereunto KR) for NPCs and events before focusing on the KRs of their parts. NPCs are 'characterized' by a set of personality traits and social roles; such that the latter indicates the social relationships, for which we note exist analogs if not equivalents in both the CiF and Needs-Based models. The paper notes that social roles are initialized WRT all other NPC's, but does not specify anything further. We note this because our work with PCG, specifically our 'Principle of Physical Stability and Contextual Consistency', provides a model by which this initialization can be qualified thereof. Events are perceived by NPCs as triggered by the world-state and characters within (both NPCs and humans); of which are 'appraised' (i.e. evaluated) by the agent. This computation outputs a 'raw emotional stimulus' from the input event; which is then transformed through an 'influence of personality' function as to factor-in the "intensity of emotions triggered in the NPC's mind" i.e. 'refined emotional effect' WRT the NPC's unique personality. This composite effect is then used to update both the NPC's emotional state and that of its social relationships as effected by the initial input event. We note that this process can be made compatible with CiF, Dialogue Agents, and Needs-Based AI (in fact, it obviously needs to be). This will be most profound with Needs-Based AI, as that model is used to simulate personality and emotion WRT physical state (i.e. being grumpy because hungry [i.e. hangry], less happy because in pain, more frustrated and confused because one is extremely tired, more excited and sociable because just drank a cup of latte, etc.) We also remind that is an equivalent function to humans processing emotional reactions; and again: because OSC is modelled from OCC (and others compatible thereof), this is a more informed mechanism than naïve methods – ergo converges more towards a more accurate simulation thereof.

Knowledge Representation

Knowledge Representations seems somewhat similar to those implemented within 'The Sims'.

Representing Actions / Attitudes / Praise

- **Actions** are {name, effect} pair where 'name' = unique ID represented as verbs in the infinitive form (e.g. 'praise', 'flirt', 'insult', 'humor'); and effect $\in[-1,+1]$ score which describes the effect of the corresponding action s.t.
  - $-1\rightarrow$ actions that have a negative impact on the character who takes this action
  - $+1\rightarrow$ actions that have a beneficial effect
  - Effect scores of actions universal to all characters (i.e. defined independent from any NPC); e.g. 'threaten' has very negative impact on person being threatened (i.e. $effect_{threaten}=-0.9$) s.t. score of 'threaten' will be −0.9 from any NPC to any NPC.

- **Praise and Action Effects / Evaluations**
  - **Praise:** Attitudes towards actions of form praise(agent,action)=$x\in[-1,1]$ which expresses that "[agent] considers the [action] to be praiseworthy if x=1 else condemnable if x=−1.
  - **Action:** Attitudes towards actions of form effect_a=$e\in[-1,1]$ | $a\in$Actions expresses that an agent considers the effect of action [a] to be beneficial to the patient thereof e=1, harmful if e=−1, and otherwise between these two extrema.
  - **Praise Effects -vs- Action Effects:** Praise evaluations are subjective to the morals and principles of the agent, whereas action evaluations are objective to the universe. This is a neat (and purposely intended) mechanism! Also the 'socio-cultural context' for which praise effects are defined tie into CiF architecture's SFKB (Social Facts Knowledge Base).

Representing Events

- **OSC Events are modeled based on Sowa Graphs** (i.e. Conceptual Graphs) wherein "actions are at the core of the world's description and linked to other entities that participate thereof using roles." However, they implement a simplified version reduced to two key roles encompassing the 'bringer of the action' (i.e. "agent") and the 'recipient of its effects' (i.e. "patient"). As we find these terms to be confusing (especially the 2nd one), we renamed them to "effector" and "affected" respectively.

- **OSC Events are thus represented by** a 4-tuple composed of 4 values: the two roles mentioned above, the action itself, and the 'Degree of Certainty' by which any NPC receiving the event (especially those outside the effector/affected) believes that the event actually occurred. We add another value of 'time' for use with certainty or inference (e.g. as Temporal Axioms/Beliefs); and note that this could represent either the time that the event occurred, the timestamp at which a particular agent receives knowledge of the event, or even both WLOG. Thus our augmentation to the tuple is $Event_e = \langle effector_e, action_e, affected_e, time_e, certainty_e \rangle$ for which we describe its parts in more detail as follows…

- **Effector:** The thing performing the action. The OSC paper limits this to only NPC agents, but we extend it to support things like groups/factions (e.g. 'Martians invaded Earth'), non-characters / groups thereof (e.g. 'Monsoon soaks Patrick'), non-character animals (e.g. 'Bee stings Waylon'), etc. It uses two special characters to represent events of which either do not know who the effector is (e.g. 'someone stole Mona Lisa') or for which an effector is unnecessary (e.g. 'Paul is

Dead'); and note that our effector domain extension is compatible with OSC because such events are handled in the same way.

- **Action:** The action of which the event represents, where action is in the set of actions $\boldsymbol{A}$.

- **Affected:** The thing which receives i.e. is-affected-by the action, such that it cannot stop it from happening; i.e. if the event is true, then the action absolutely took place such that the affected received its full effects.

- **Time:** As mentioned, the OSC model does not include a temporal representation into the 4-tuple; though it notes that the event's timestamp is received and used in a version of the emotional intensity computation step with the decay-over-time function, as to simulate things like 'calming down after hearing about a disaster, false alarm, getting less surprised in the minutes and hours after the start of one's surprise party, etc.

- **Degree of Certainty (DC):** Degree by which agent receiving the event is confident that it actually took place; and is a surprisingly powerful value. It is expressed in the range $[\mathbf{0}, \cdots, \mathbf{1}]$ where:
  - $\boldsymbol{DC} = \mathbf{1} \rightarrow$ Agent is absolutely sure that the event occurred as reported.
  - $\boldsymbol{DC} = \mathbf{0} \rightarrow$ Agent once thought event was possible, now knows it did NOT happen.
  - $\boldsymbol{DC} = (\mathbf{0}, \cdots, \mathbf{1}) \rightarrow$ Agent believes event is more or less possible to have happened.

- **Who provides Events to NPC?**
  - The paper provides two methods for how NPCs 'hear about' events: being informed by some other NPC, or 'connecting the dots given some clues and its knowledge/memory. Implementation-wise: this means character dialogue (especially NPC-to-NPC) and logical inference via the NPC's KB (hey – didn't we have a unit on this?) We add two more methods vis-à-vis existing mechanics used for other games: direct observation of the event 'in the world' (e.g. witnessing an incoming swarm of zombies or aliens abducting one's friend), and indirectly being told about an event via news broadcast, newspaper, giant signs, etc. within the world. We reserve discussion on implementation of anything besides character dialogue (though note that initial research was done for this paper on streamlined KR/KB inference models, and the implementation of 'broadcasting news' to NPCs is a [relatively] straightforward one with well-implemented precedent).

- **How is Degree of Certainty Computed?**
  - The paper offers the same thoughts as with the previous question.
  - For determining DC on an event received from a fellow NPC (via a dialogue exchange), it suggests using the agent's reasoning engine to do logical inference towards judging how trustworthy the interlocutor is; but does not discuss any further implementation details towards how to implement this. We, however, do have a method within our composite architecture; as the CiF model has a simple yet powerful means by which to represent and infer about qualities/traits of another character!
    - We also note what we think the authors also observed and were also hesitant to detail/speculate further: the combination of receiving news from other NPCs as well as the history of how much 'gossip' ended up being correct (xor not) steps us into an even greater dimension of interpersonal dynamics vis-à-vis how an NPC updates their viewpoints on the trustworthiness of specific fellow NPCs and groups thereof. This therefore could realize everything from 'high school rumor-

mill drama' to a nation of people growing distrustful of the propaganda, news, and otherwise claims made by their government (bringing us close to the edge of even novel models and experiments thereof in sociocultural, interpersonal, and communications/marketing simulation!)

- For determining DC on an inferred event: this foremost clearly requires an NPC having a reason for making the inference; which points to the agent's realization of Beliefs, Goals, and Desires (Clay told me BDI was hyper-relevant!) Regarding implementation, we must likewise restrain from additional details due to the scope of this paper/project; but will note that we have done initial research into one promising solution method in GOAP (Goal-Oriented Action Planning).
- Lastly, for observed events: this comes down to how much the agent trusts their own percepts. This leads us into another area of major AI research in and of itself, and we will in brevity resolve DC for perceived events by naively setting the value to 1 i.e. implying 100% confidence that the event occurred.

- **A Final Note on KB Composition of Complex Events:** The paper makes a good point to introduce a possible disconnect between the Concept Graph representation used versus how complex events would be communicated in a script via natural language. For example, consider the narrative "Francis is alone cornered in an alleyway fighting off a zombie who tackled him to the floor and just bit him" from the perspective of Francis, the zombie, and anyone who observes / hears about this encounter. We would need to compose the following 'event atoms' as to tell this story; but before we do, we will note that implementation wise (i.e. at-runtime, online) – there exist both procedural and ML methods which could decompose complex events into 'atomic' form. That said:
  - $\langle \emptyset, aloneAndSurrounded, Francis, t, 1.0 \rangle$
  - $\langle Francis, fighting, Zombie\#51631, t + i, 1.0 \rangle$
  - $\langle Zombie\#51631, fighting, Francis, t + i, 1.0 \rangle$
  - $\langle Zombie\#51631, tacklesToFloor, Francis,\ t + 2i, 1.0 \rangle$
  - $\langle Francis,\ physicallyResisting, Zombie\#51631, t + 2i, 1.0 \rangle$
  - $\langle Zombie\#51631, Bites, Francis,\ t + 3i, 1.0 \rangle$
  - $\langle \emptyset, zombified, Francis, t + z, 0.95 \rangle$
  - $\langle Zombie\#51631, kills, Francis, t + k, 0.55 \rangle$
  - $\langle ZombiePlague, kills, Francis, t + z, 0.95 \rangle$
  - **Notes:** Notice how we represent concurrently possible events here. That Francis was bitten means he thinks he will be zombified by some future time inferred within his KB, and that he will consequently die [well, un-die actually] at that same time. But he also thinks he has a 55% chance to die during the fight itself (i.e. having his brain eaten) at a much shorter time. Just as we would contemplate and feel emotions for an event that suffers an immediate and future consequence: by our model – so will poor Francis!

- **Mathematical Expressions for KB elements:**
  - **Domains:** $\{\boldsymbol{A}, \boldsymbol{O}, \boldsymbol{N}\}$ where $\boldsymbol{A} \rightarrow$ **Set of all actions**, $\boldsymbol{O} \rightarrow$ **Set of all objects** (i.e. anything not an NPC or human, and $\boldsymbol{P} =$ **Set of all characters** (for which we note that the paper doesn't seem to include humans into this set, but we do)

  - **Actions:** $\forall \boldsymbol{a} \in \boldsymbol{A} : \boldsymbol{a} \ni \{\boldsymbol{name}, \boldsymbol{effect}\}$ where $\boldsymbol{effect} \in [-\mathbf{1.0}, \cdots, \mathbf{1.0}]$ and $\boldsymbol{name} \rightarrow \boldsymbol{ID}$ and is represented as a verb in the infinitive *(seemingly as a String for the paper, though ints, enums, dicts, etc. could provide quicker identifier in code, WLOG)*.

- **Attitudes:** $\boldsymbol{AT} = \boldsymbol{N} \times (\boldsymbol{O} \cup \boldsymbol{N})$ we express as implicit $|\boldsymbol{N}| \times (|\boldsymbol{O}| + |\boldsymbol{N}|)$ matrix where $(\boldsymbol{a_{i,j \cup k}} \in \boldsymbol{AT})$, $\boldsymbol{a_{i,j \cup k}} \in [-\mathbf{1} \cdots \mathbf{1}]$, $(\boldsymbol{i} \to \boldsymbol{n_i} \wedge \boldsymbol{j} \to \boldsymbol{n_j} | \boldsymbol{n_i}, \boldsymbol{n_j} \in \boldsymbol{N} \wedge \boldsymbol{i} \neq \boldsymbol{j})$, and $(\boldsymbol{k} \to \boldsymbol{o_k} \in \boldsymbol{O})$. This form expresses that each row corresponds to a feature vector of length $|\boldsymbol{O}| + |\boldsymbol{N} - \mathbf{1}|$ whose column values represent the attitude of exactly one agent $\boldsymbol{i}$ towards all objects $\boldsymbol{k} \in \boldsymbol{O}$ and all other characters sans itself $\boldsymbol{j} \in (\boldsymbol{N} - \boldsymbol{n_i})$. We also added that latter part as to assert the agent in question has no attitude value for itself, as such would (or could) be encompassed by its 'self-esteem' value.

- **Praise:** $\boldsymbol{PR} = \boldsymbol{P} \times \boldsymbol{A}$ which can be expressed as a similar matrix of dims $|\boldsymbol{P}| \times |\boldsymbol{A}|$ where $\boldsymbol{PR}(\boldsymbol{n}, \boldsymbol{a}) \in [-\mathbf{1} \cdots \mathbf{1}]$ represents attitudes towards actions as defined above.

Representing Emotions

- **Intro:** As discussed above, emotions are modelled via simplification of OCC model. In terms of their consequent implementation: the OSC model considers emotions WRT to their "triggering conditions"; for which several good example are discussed which we replicate as follows:
  - $(joy \oplus distress)$ caused by $(desirable \oplus undesirable)$ event;
  - $(hope \oplus fear)$ caused by expectation of $(desirable \oplus undesirable)$ event;
  - $(relief \oplus disappointment)$ via non-occurrence of expected $(undesirable \oplus desirable)$ event;
  - $(pride \oplus shame)$ following $(praiseworthy \oplus blameworthy)$ action done by the agent;
  - $(admiration \oplus anger)$ via $(praiseworthy \oplus blameworthy)$ action done by another agent.
  - **NOTE:** Don't forget to copy over Figure 2 [PDF 10]

- This OSC has 10 emotions: {joy, distress, hope, fear, relief, disappointment, pride, shame, admiration, anger}

- Emotions are represented via variables of values in the range [0, 1]. The set of emotions of NPC i at a time t is then represented by a vector $e_i(t) \in ([0,1])^{10}$ in which e_i^j give the intensity of the emotion $j \ni$(joy, distress, etc).

- 3 different emotional vectors defined to represent:
  - Emotions triggered by an event (noted e_di(evt) for NPC i for event evt): these emotions depend both on the emotional stimulus associated with the event and the NPC's personality;
  - the emotional state of an NPC (noted e_ei(t) for NPC i at time t): these emotions represent the set of emotions felt at a given time by an NPC. They are updated according to triggered emotions and decrease naturally through time.
  - the emotions expressed by character i at time t (noted e_expi(t)) : a character may decide to express emotions different from its felt emotions [40].
  - OSC focuses on first two i.e. NPCs' triggered emotions and emotional state, but could speak about Streamline work as external partial realization of 3rd point outside of scope but for which I have some practical dev experience.
- WRT the 'raw to refined emotion' steps: This encompasses producing a raw emotional stimulus from the event relative to an NPC with a neutral personality (as for the emotional intensity and other steps to transform this value WRT the NPC's individual personality).

- **TODO: IOU MY EXPLANATION OF FIGURE 2 (PAGE 10) ###**

Representing Emotional Intensity Speaking of which…

- OCC model positively correlates intensity of emotions {joy, distress, relief, disappointment, hope, fear} to event in terms of its degree of desirability/undesirability and probability of occurrence (in the case of hope and fear); and analogous for {pride, shame, admiration, anger} WRT praiseworthiness/condemnation.
- OSC realizes degree of desirability i.e. intensity of its associated emotions based on effect of action, attitude of recipient agent i toward affected agent, and probability of event occurrence i.e. DC value. Degree of praiseworthiness i.e. intensity of its associated emotions analogously handed via the $praise$ function.
- **TODO: IOU also MY EXPLANATION OF FIGURE 3 (PAGE 11) ###**

Influence of Personality on Emotions:

- Paper points out social science / psychology research shows that personality influences both emotions and corresponding behaviors. While it notes that the OSC model does not cover influence of personality on behavioral aspects such as action selection and expression of emotions, we can and will briefly.

- WRT action selection: we will first also include task, schedule, and goal planning based on emotions; and note that architectures/methods we've brought up throughout this paper such as Procedural Reactive Agents (realization of Belief-Desire-Intention Model), Black-Board Design Pattern, and Goal-Oriented Action Planning (implemented with Behavior Trees and Decision Networks) can in theory (and from our dev experience vis-à-vis Utility Behavior Trees) implement this feature. However, as with paper: this is kind of out of scope so we'll stop here.

- WRT emotional expression: Project Streamline, 'Nuff Said.

- OSC Model does however implement influence of personality on emotion via three 'impacts' thereupon and the mechanisms which realize them. They are as follows, as discussed below:
    - Emotions triggered by event
    - Temporal decay of intensity
    - Emotion intensity threshold

Impact - Emotions triggered by event:

- Here is where the 'Big-5' Personality Traits model comes into play. The paper provides some rules for how each of the traits affects the emotional stimulus that a NPC has for an event. To recap: this is to support the idea that NPCs of two distinct personality types can process the same event with different emotions if not degrees thereof. The rules are as follows:
    - Extroverts feel higher intensity of positive emotions than otherwise, with no special influence on negative emotions
    - Neurotics feel negative emotions with a higher intensity than otherwise, with no special influence on positive emotions.

- "Agreeable agents feel empathy-oriented emotions (e.g. distress for others) with a higher intensity and negative emotions directed toward the interlocutor (e.g. resentment) with a lower intensity."
- We note from our earlier discussion on personality modelling that neither OSC nor our report provides substantial detail on Conscientiousness and Open-Mindedness, though we did provide some high level theoretical discussion and speculation thereof.

- Furthermore, OSC did not implement Agreeableness, because they did not represent emotions involving interlocutors (i.e. relationships and feelings about other characters); as such would require modelling the "evaluation of emotional impacts of events on others [and their] preferences". However, this is another case of where OSC and CiF couple brilliantly – as CiF does model a kind of analog which can at least partially fulfil this purpose. The paper did, however, speak for two pairs of interpersonal-based emotions:
  - $(happyFor \oplus sorryFor)$ occurs when an agent finds out about a $(desirable \oplus undesirable)$ event that occurred to someone they like;
  - $(gloating \oplus resentment)$ occurs when an agent finds out about a $(desirable \oplus undesirable)$ event that occurred to someone they dislike;

- That said: personality represented by a vector whose values are in the range $[-1,1]$ where each value corresponds to a personality trait in the affirmative such that its negative counterpart (e.g. value of 'extroversion' would imply 'introvert' if negative, neutral if zero, 'extrovert' if positive). As OSC only uses 2 traits and we support a third: our personality representation would thus be a 3-tuple of the form $PS = \langle p^{extrovert}, p^{neurotic}, p^{agreeable} \rangle \mid \forall\, p^x \in PS,\ p^x = [-1,1]$

- As to the impact itself: Directly computed in emotional stimulus model and personality trait values. "For a given value of emotional stimulus, the more the NPC is extrovert, the higher the triggered joy, hope, pride and relief emotion values will be. Conversely, the more the NPC is neurotic, the higher the values of distress, fear, shame and deception will be. This variation in the emotion intensity corresponds to a multiplicative factor on the emotional stimulus"

Impact - Temporal decay of intensity:

- "The decrease function characterizes temporal evolution of intensity of an NPC's emotion when no emotional stimulus occurs. The intensity of emotions in the NPC emotional state naturally decreases when no emotion is triggered. The decrease of each emotion is defined by a monotonous function that depends a priori on personality and that tends toward neutral state. For instance, negative emotions are often considered as more prevailing than positive emotions. In our model, each component of the emotional state's vector will be attached to a specific decrease value, whose expression can depend on the personality."

Impact - Emotion intensity threshold:

- **They define three thresholds** of which further implementation details are out of scope for OSC (and mostly our report) as they encompass affecting how the agent visibly expresses emotions and the functionality of their reasoning i.e. decision-making / inference system.

- **Base Intensity Threshold:** This sets a baseline at some value $(0,1]$ s.t. if the emotional stimulus is below that value: it is effectively masked to zero as to have absolutely no effect on either the behaviors the agent makes, nor the gestures/expressions it displays.

- **Emotion Activation Threshold:** This sets a baseline at some value $(0,1]$ s.t. the behavior/task system will recognize the emotional stimulus only if it's above the baseline. The paper rightly notes that this sets up a discrete (if not binary) thus symbolic means by which the other systems can utilize the value. That is: a (Behavior Tree, Decision Tree, GOAP-Graph, etc. WLOG) node can route traversal of an action query based on this value much easier than with a continuous range, and without having to support any other computation sans its own qualification method.

- **Saturation Threshold:** Sets up a baseline at some value $(0,1]$ s.t. the agent's rational decision-making system will be adversely affected if the emotional stimulus is above this baseline. That is: value for which the emotion is strong enough for the agent to act irrationally.
    - Example scenario: An very frightened NPC running from ghosts in a haunted mansion will suffer sub-optimal planning, pathing, etc. as to irrationally decide on and compose the plan 'RunAway!' which acts as a combination dynamic pathing/steering agent as to constantly direct the character in the opposite direction of the ghosts. If they were not as frightened, or at least could think rationally: two alternate plans are possible which can eliminate the danger outright versus deferring it a-la 'RunAway!' The plan 'GTFO!' [acronym beyond the scope of this paper] is also a dynamic pathing/steering behavior; except it involves bringing the agent to the closest ground floor exit and off the property to escape, though it could have been rejected because a correct path and solution might have brought the agent close to or in the direction of a ghost. The plan 'GhostBust!' is more direct: and involves pathing to a nearby room to equip and use the agent's trusty UAC Series-3 Plasma Gun which is very capable of turning the ghosts into ectoplasm.

Representing Social Dynamics / Relationships:

- OSC represents the relationship of NPC p to NPC q at time t from p's POV with a 4-tuple as follows: $socialRelation_{p,q}(t) = \langle liking, dominance, familarity, solidarity \rangle$. We augment this representation into an equivalent compact KB-oriented form $SR(p,q,t) = \langle d^L, d^D, d^F, d^S \rangle$. In any case: the values $\{d^L, d^D, d^F, d^S\}$ are in the range $[-1,1]$ and represent the degrees of liking, dominance, familiarity, and solidarity respectively, at time t; where the meanings of these variables correspond directly to their definitions earlier in this section.

- The OSC model further defines social relationships WRT 'complementary social roles' e.g. 'student/teacher', 'employee/manager', 'child/parent', 'man/woman', 'doctor/patient' etc. s.t. there exists a large set of possible roles and any complementary pairs could be formed thereof, e.g. 'student/parent'. Specifically: they say "that research shows social variables depend not only on individuals but also the roles between themselves". Thus, whereas we discussed earlier that the initial values of social relations were 'Contextually Consistent', OSC actually offers a compatible mechanism for how to procedurally initialize these values via social roles.

- That also being said: the utilization of social roles need not be limited to initialization of social variable vals. CiF and Dialogue-Agents has somewhat of an analog such that if all were merged: CiF would continue to use social roles as perhaps a context for speech options / action options, while Dialogue system can populate player responses or NPC procedural dialogue via them WRT

the role (e.g. speaking politely to police officer, using simpler language responses when talking with children, etc.)

- They first provide representation notation: Given the set of roles R and two roles $(r_x, r_y) \in R^2$, for which we believe $R^2 = SR$ as the Cartesian Product of R as to compose a $|R| \times |R|$ lookup matrix whose values are accessible via $SR(p, q, t)$ s.t. such values encompass the social role of p relative to q at time t, i.e. $socialRole^t_{p|q}$. The social role value is a 4-tuple $\langle d^L, d^D, d^F, d^S \rangle$ whose values are as defined above; except if the roles are NOT complementary e.g. 'child/patient', then the tuple is a null vector or has all values set to zero (WLOG).

- Multiple Roles: The paper notes that "a character may endow several roles at the same time during an interaction with another character", and that the values therein need to be combined; for which the OSC model uses a naïve average value between the non-null roles.

- Social Distance: An ironic phrase nowadays, but OSC defines it as the solidarity between two persons, and notes that it "depends on the solidarity of each of them with others in the local social universe. For instance, two French individuals in a foreign country might be more in solidarity than in France. This phenomenon is explained by the fact that "wherever one is, all members of the social local universe must be assigned a place across the entire scale of social distance" [5]. Then, in our model, the value of role solidarity equals the value of solidarity of the role r_i in relation to the role r_j divided by the maximum value of solidarity that the character with role r_i has with the others characters of its environment. Consequently: all agents of the social local universe are assigned across the entire scale of solidarity."

Emotions-based Social Dynamics (Largely directly quoted!)

- AKA: You init'd the social variables via role-to-role matrix at initialization, now how do those values change over time? "during an interaction, a person's emotions and those of his interlocutor may lead to a change in their social relations, for which we analyze separately the dynamics of each variable of a social relation of NPC i during an interaction with another character j according to their emotions. If no event occurred, we suppose that the social relation does not evolve. In the contrary case, the intensities of the emotions triggered by the event are used to update the social variables:"

1. Liking: The degree of liking one has for another depends on the valence (positive versus negative) of the emotions induced by the latter. Thus, we model that a positive (resp. negative) emotion of i caused by j induces an increase (resp. a decrease) in the degree of liking i has for j.

- **TODO: IOU also MY EXPLANATION OF FIGURE 4 (PAGE 14) ###**

2. Dominance: Pride and anger reflect a domination whereas shame, distress, and admiration correspond to an inferior status. Consequently, we can model that emotion of pride or anger of i caused by j induces an increase in the dominance value that i thought to have on j. Conversely, an emotion of fear, distress, admiration, or shame of i caused by j infers a decrease in the dominance i thought to have on j. Moreover, in [26], expression of distress or fear reflects a low value of dominance. Finally, some types of emotions expressed by someone affect the

dominance value of the person who perceives it. We can model that the expression by j of an emotion of fear or distress induces an increase in the dominance of i. Consequently, if the event triggers fear or distress for i and j with the same intensity the dominance of i on j won't change.

- **TODO: IOU also MY EXPLANATION OF FIGURE 5 (PAGE 15) ###**

3. Solidarity. Negative emotions caused by another person leads to a decrease in the solidarity value whereas the triggering of positive emotions does not modify this value. We model that a negative emotion of i caused by j induces a decrease in the value of solidarity that i thought to have with j (Figure 6). Moreover, [25] shows that the expressed emotions reflect a person's mental states and then indirectly his goals, beliefs, expectations, plans, etc. Consequently two individuals expressing similar emotions in reaction to a same situation should feel more in solidarity. Inversely, expression of opposite emotions may lead to a decrease of solidarity. Then, in our model, an incongruence (resp. congruence) may lead to a decrease (resp. increase) in solidarity. If the triggered emotion of i is joy or hope (and is not caused by j) and j expressed emotion of the same type, the solidarity increases. Further: research shows solidarity influences liking; and that it appears one likes more similar persons [49]. Then, we can suppose an increase in solidarity of i with j induces an increase in the degree of liking i has for j.

- **TODO: IOU also MY EXPLANATION OF FIGURE 6 (PAGE 15) ###**

4. Familiarity. In the literature, emotions seem to not have a direct impact on the familiarity (i.e. on the degree of confidentiality of the information transmitted by a person). However, research shows that one confides more in another when the former likes the latter. Therefore, we model that the more the character likes another one the more it will transfer a confidential information to it. In other words, the familiarity is indirectly connected to the liking variable. Indeed, an increase in the confidentiality of the information transferred leads to an increase in the familiarity value. This mechanism depending on the implementation of the NPC's behavior is not presented in this paper.

Implementation

Basically – the paper continues with mathematical definitions of functions used with their implementation of the OSC model. We leave discussion of these definitions out of this report, as we've captured [more than enough] detail on this model WRT the scope of this report; but note that we would refer back to this part to attempt a small experimental implementation of our own, as we're actually planning on maybe doing over the summer – who knows?

Conclusion

In the 'Conclusions and Future Work' section, the paper states a similar idea to ours in terms of a more composite realization via other external systems: "the benefits of our enriched modeling will indeed appear more clearly when its integration in a full model with action selection and behavior control will be completed and tested". It lastly mentions ideas for expanding their emotion representation. Where it currently directly connects emotions to resulting behavior, one of their other Psychological Models, PAD, utilizes a "dimensional emotion representation ... using pleasure, arousal, and dominance dimensions ... to facilitate the connection between emotions and behavior." And furthermore, they

envision NPCs that are able to "take into account the emotional and social impact of their actions so as to guide their strategy for dialog, and more generally, for behavior." Towards both of these concepts: we note that the CiF Model discussed in the next section provides at least a basic realization of these ideas.