

PROJECT FÈA

ILU ILÚVATAR EN KÁRE ELDAIN A FÍRMOIN

“Project Fëa: Simulating Personality, Emotions, Social Relations, and Physical/Mental Needs for autonomous Non-Player Characters within video game / computer simulation worlds”

Steven Eiselen | seiselen24@email.arizona.edu | Final Version

Table of Contents

Foreword.....	2
Section 1 – Introduction	2
Overview / Motivation	2
Applications / Aspirations	3
Objectives / Approaches	6
Section 2 – OSC Model	7
OSC Model Introduction.....	7
Attitudes and Praise Model.....	9
Personality Model	9
Emotion Model.....	10
Social Relations Model	13
Event Representation Model	15
Section 3 – CiF Model	17
CiF Model Introduction	17
Background Models / Analogs	18
CiF Architecture Overview.....	19
NPC Character Representation	20
Social State Representation	21
Social Exchange Representation	24
Section 4 – Dialogue Agents Model.....	26
Dialogue Agents Introduction	26
Dialogue Types and Elements	27
Agent-Based Dialog / Turn-Taking.....	28
Affective Communication.....	29
Section 5 – Needs-Based AI Model.....	30
Needs-Based AI Introduction	30
Needs and Types thereof	31
Implementation within AI system	32
Section 6 - Conclusion.....	33
Section 7 - Bibliography	34

Foreword

Throughout my Computer Science studies, there have been a few opportunities to make progress on the quest that compelled me to enter this field back in 2012. It was not for career purposes, nor because I am naturally gifted in mathematics, programming, etc., but because I became enthralled towards determining how to build the most realistic simulated worlds possible and provide the most immersive interfaces by which humans could explore them. My work with immersive VR/visualization has somewhat addressed the latter goal, while the other goal of realizing simulations whose worlds are as realistic as possible and whose inhabitants within are as 'alive' as possible encompasses a sub-quest I've named 'The Music of the Ainur' after the Creation Story of the Tolkien Universe; while naming its two subparts within Æa and Fëa.

Æa is the Quenya word for 'Created World', thus its analog in my research is 'Project Genesis' which encompasses my studies and experiments into the use of Procedural Content Generation (PCG) for the autonomous creation of simulation worlds. Fëa is the Quenya word for 'soul' or 'spirit', and was the project codename given to this research paper. Where my work on Project gave me a chance to advance my understanding of how worlds were generated beyond the glimpses I was previously only able to gain between the cracks of curriculum coursework, I needed a similar opportunity explore how to bring 'life' to NPCs; which is what this term project has provided. Therefore: while there remain many topics yet to research, projects to build, and experiments to implement (surely a lifetime's worth) – Project Fëa, of which this report is its artifact, stands as a fitting conclusion towards finally exploring this missing half!

Thank You Clay for all of your support along this journey!

Section 1 – Introduction

Overview / Motivation

"Ever since the first computers, there have always been ghosts in the machine. Random segments of code that have grouped together to form unexpected protocols. Unanticipated, these free radicals engender questions of free will, creativity, and even the nature of what we might call the soul...When does a perceptual schematic become consciousness? When does an [inference] engine become the search for truth? When does a personality simulation become the bitter mote...of a soul?"

- Dr. Alfred Lanning, Director of Research at U.S. Robotics [from the film *I, Robot* (2004)]

This report discusses the simulation of Personality, Emotions, Social Relations, and the Physical/Mental needs of autonomous Non-Player Characters (hereafter NPCs) within computer simulated worlds (e.g. video games); as well as some discussion on the Knowledge Base, Inference Reasoning, Goal Planning, and Action Tasking implications thereof. Personality Simulation, roughly speaking, encompasses the implementation of Personality Traits, Attitudes (a.k.a. Preferences), and Emotions for NPCs based on their Needs, Goals, and Desires. Social Simulation involves a likewise realization towards NPC's that are able to interact with other characters, form and grow relationships between themselves over time, and otherwise participate in and emotionally react to the social dynamics of the societies within their world. Regarding the latter: social simulation also applies to their relationships with humans (known as 'Player Characters', hereafter PCs). Beyond the deeper significance of this topic as discussed in the Foreword, we have found personality and social simulation to be one of the most fascinating [sub]topics of artificial intelligence and immersive features of video games; intrigued by the capability of NPC's able to express their own personalities and preferences in both the tasks they chose to partake in, the characters they chose to associate with, the dynamically changing emotions they felt in response to all kinds of internal needs and external stimuli, and the corresponding behaviors they exhibit in response thereof.

Applications / Aspirations

As with all topics involving future tech: Star Trek plays a role; specifically: The Holodeck. This technology encompasses a hyper-advanced Virtual Reality Simulation such that humans / other races can physically interact with it and the agents who exist within. These characters, NPCs by definition (whose realism all but qualifies them as real living beings), are often the stars of the show. Specifically, via interacting with 'real' guests as to guide quests, contribute to the narrative of a scenario, provide training and scientific visualization, etc.; and such utilization of NPCs is exactly what we see in today's video games and other applications, as discussed below. In fact, one of our reference models, CiF, defines a social relationship and dynamics model oriented on theatre, based on props and performances from a character to other

characters, (emotional) exhibitions thereof, etc. Another model, OSC, provides an great qualification to this idea, stating: *“The success of a video game...relies greatly on its ability to give a feeling of immersion [enabled by] willing suspension of disbelief”*; where the *“degree of consistency between elements of the world (events, spaces, NPCs, etc.)”* is a necessary condition whose sufficiency *“relies significantly on [NPC] behaviors”*. They go on to say that NPC’s are thus *“...key elements to the narrative [such that] their credibility is key to the player’s perception that they are observing truly living beings.”* The qualification of consistency, whose metric is defined as a degree of credibility, is discussed in Section 2.

Beyond video game entertainment, we note a second major application in pedagogy via experiential immersion. As the Holodeck characters in Star Trek TNG helped bring Noir Detective stories in 1940’s San Francisco and ancient Klingon rituals to life: our work on various iterations of Virtual Harlem helped bring the 1920’s Harlem Renaissance to life; despite the fact that we weren’t even able to implement NPCs but only humanoid characters on animation loops. Still – having those background characters in the scenes increased immersion immensely compared with scenes with no characters; a term we’ve dubbed ‘The Langoliers Effect’ (rec). Another example is the USC Center for Creative Technologies’ ‘Virtual Human Program’, where interactive NPCs with Knowledge Bases (hereafter KB’s), some of which are even capable of speech-to-speech communication with humans, are used for ongoing experiments with psychotherapy, conflict resolution simulation for the Armed Forces, and others. Indeed: the use of intelligent NPCs in VR for training purposes has a big example right here at UArizona: Leonard Brown’s “Harry’s Hard Choices” which is used for mining safety training and emergency management drills.

Two other ‘classic’ applications warrant discussion, as despite realizing different kinds of agents, they are effectively analogous to NPCs of high intelligence, personality, thought, engagement, etc. except that they are each embodied; i.e. exist ‘independently’ within their own virtual or physical body, versus existing

‘dependent’ to some simulation world. The first type are Virtual Assistants/Agents; for which we note recent advancements made with Apple Siri and Microsoft Cortana, and previous work such as the [notorious] attempts by Microsoft to integrate virtual assistants into their products (here’s thinking of you, Clippy!) The second application are humanoid robots i.e. androids. Robots with awareness and personhood have been depicted in film by Bicentennial Man (1999), I-Robot (2004), Ex Machina (2014), and even the classic Star Trek TNG character of Data (and Lore); as well as both since the conception of robots and automatons, and as far back as human antiquity (e.g. Golems of Jewish tradition, Talos of Greek tradition). And as seen in many of the [non-apocalyptic] depictions of such robots: their purpose is largely the same as virtual assistants; except they are able to physically interact within our world.

We discuss a final application, yet one that is the most...existential: modelling complex simulations of worlds inhabited by virtual humans for strictly experimental and observational purposes. To introduce this, we invoke the other major Sci-Fi series that always gets brought up when discussing Virtual Reality and NPCs in the same sentence: ‘The Matrix’. Simulation Hypothesis involves a probabilistic argument for the likelihood of our universe being a simulation (for which we must be witty in noting this would, by definition, make us all NPCs!) The specifics aside, we invoke it for its premise: that a future civilization interested in observing their past may find it more practical or even fully satisfiable to implement a virtual simulation than to implement time travel. And indeed: Star Trek’s Holodeck, Virtual Harlem, and even some simulation-oriented video games are precedent for replicating a virtual representation of the past for purposes of science, engineering, exploration, and entertainment. Indeed: a variant for one of our projects while employed at CDH involved a simulation of Ancient Roman businesses wherein the client had significant historical reference for elements which were recorded (i.e. inventories, prices, contracts, even insurance policies), but did not have any reference for how the day-to-day operation of business went. Thus, we proposed modelling a simulation from the known data as to ‘run it and see what happens’.

Objectives / Approaches

Having introduced the fundamental ideas/applications, we discuss the corresponding problem definition based on objectives/approaches towards realizing NPCs with the desired capabilities. The perspective by which we 'theme' our discussion of the models we studied is of developing our own concept model that realizes our topic, i.e. simulating robust and dynamic personality, emotions, social relationships, physical needs, and mental needs for NPCs; while providing some discussion on the Knowledge Base, Inference Reasoning, Goal Planning, and Action Tasking implications thereof. We call this concept model Fëa, same as the project. And while we researched several excellent and promising methods for the latter part (i.e. KBs, behaviors, etc.): the size and scale of the first part (i.e. personality, social relations, etc.) proved to be WAY too much material for this project, so we needed to significantly reduce its scope to only some peripheral discussion where possible. We researched a several personality and social dynamics models (from a surprisingly limited population); discovering four of which are discussed in this paper. Thus in answer to the question of what the 'state of the art' is for personality and social simulation for NPCs as of 2015, we would argue that they're in this paper; sans proprietary methods, robotics applications, and insufficient 'pre-scripted' i.e. 'handmade' approaches. The merit of the models we studied can be more specifically qualified: Needs-Based AI originates from The Sims, OSC and CiF use major cognitive/social psychology models, and Dialogue Agents is a summary of methods used for major video games. While we have interesting ideas for why this is the case, such discussion is out of scope. Lastly, it turned out that each of our models has enough overlap in terms of Knowledge Representation (hereafter KR) and other features as for them to be more compatible with each other than expected; while also each having unique features desired for Fëa as for the combination of these models to each provide substantial contribution. In other words: the means by which these models complement each other is delightfully substantial!

Section 2 – OSC Model

OSC Model Introduction

We begin with a synopsis of the ‘Ochs Sabouret Corruble’ Model (hereafter OSC) which we named after its principal authors, encompasses a *“model of personality, emotions and social relations aiming at the improvement of NPC credibility in video games”*. As discussed in the previous section: credibility is an important idea for this model and NPC AI in general; as it speaks for how immersive a simulation is WRT the sum of all its components towards effecting a ‘willing suspension of disbelief’. Our research into Procedural Content Generation yielded a similar concept which we define within the “Principle of Physical Stability and Contextual Consistency” regarding the generation of simulated worlds. Thus, in a sense, this report could be seen as a corollary to Project Genesis; especially as OSC’s defines credibility with two key dimensions of ‘Behavioral Consistency’, which function analogous such that they could satisfy Contextual Consistency in our Genesis model for 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 present world. 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 with others (i.e. having an existing opinion about a character); of which together compose the current 'attitude' an NPC has with another character. 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.

With the Background Models identified and introduced: 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 thereof. NPCs are 'characterized' by a set of personality traits and social roles; such that the latter indicates social relationships, for which we note exist analogs if not equivalents in both the CiF and Needs-Based models. Events are perceived by NPCs as triggered by the world-state and characters within (NPCs and humans) which are 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.

Attitudes and Praise Model

Actions are represented as a $\{name, effect\}$ pair where *name* is a unique ID represented as verbs in the infinitive form (e.g. 'praise', 'flirt', 'insult', 'humor') and effect is a score describing the corresponding action's effect in the range $[-1, +1]$ where -1 implies actions of negative effect upon the recipient and $+1$ implies those of positive effect. Effect scores are universal i.e. shared between all characters, e.g. 'threaten' might have an effect score of -0.9 , such that any agent processing the action would see it very negatively. Praise is a somewhat similar variable encompassing attitudes towards actions and is also the range $[-1, +1]$ such that an agent considers some action to be praiseworthy if its value is $+1$, condemnable if -1 , and somewhere in between otherwise. Praise evaluations are subjective to the morals and principles of the agent, whereas action evaluations are objective to the universe. This is a very neat implementation, and analogous to the Social Facts KB of the CiF model discussed in Section 3.

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. It discusses 'category-based models' (e.g. personalities defined atomically as 'good', 'bad', etc.) and 'attribute approaches' (e.g. 'charisma' value, typically in range $[1,10]$ is used to compute probability that a PC can persuade an NPC to provide information, switch factions, etc.) These methods compose most of the implementations seen in games (even today); but suffer the problems of being mostly to fully pre-scripted, thus making it unsatisfactory for our desires. 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). It is well-known in cognitive psychology, and utilizes a “*detailed enumeration of behavior categories attached to each personality trait*”; of which such trait values are used to help define emotions and attitudes for NPCs. OSC asserts that this mechanism 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, alongside an equivalent analogy via Wikipedia: (*Extrovert/Introvert*) implies (*outgoing / solitary*), (*Neurotic/Stable*) implies (*nervous/secure*), (*Agreeable/Disagreeable*) implies (*compassionate/detached*), (*Conscientious/Neglect*) implies (*organized/careless*), and *Open-Mindedness* implies (*curious/cautious*). Personality is represented by a vector where each value is in range $[-1,1]$ and corresponds to a personality trait s.t. our representation would be a 3-tuple of the form $PS = \langle p^{extrovert}, p^{neurotic}, p^{agreeable} \rangle \mid \forall p^x \in PS, p^x = [-1,1]$. OSC only realizes 'Neurotic' and 'Extrovert' as the others depend on external agents' emotions or perception thereof, else other out-of-scope factors. Fortunately: CiF discussed in the next section can fill this gap, thus our method could certainly realize Agreeableness, likely some degree of Conscientiousness, possibly even Open-Mindedness!

Emotion Model

The Emotion Model encompasses when certain emotions should appear and in what degree/intensity: as to implement how they are displayed (e.g. dialogue selection, facial expression, body gesture animation); and where plans, behaviors, and actions can all result from them. It uses a small mechanism of behavioral mathematics to 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 very scared negatively impacts rational decision-making’. OSC introduces 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 use of the 'Ortony Clore Collins' Model (hereafter OCC) which models processing emotions via based on evaluating 3 things: consequences of an event on the agent's goals, actions performed by the agent, and objects in the environment. OSC adapts OCC based on its idea that *"attitudes of NPCs towards actions, objects and other characters of the environment."* Further: OCC proposes a *"set of simple and easily implemented rules"* to indicate which emotions are triggered and in which situations; as 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"*, and *"emotions are triggered by agent's belief that probability to complete one of its goals has been modified."* CiF uses 10 emotions: $\{joy, distress, hope, fear, relief, disappointment, pride, shame, admiration, anger\}$ Further, OSC realizes emotions via "triggering conditions" for which several good example are shown as follows: $(joy \oplus distress)$ caused by $(desirable \oplus undesirable)$ events, $(hope \oplus fear)$ caused by expecting $(desirable \oplus undesirable)$ events, $(relief \oplus disappointment)$ via non-occurrence of expected $(undesirable \oplus desirable)$ events, $(pride \oplus shame)$ following $(praiseworthy \oplus blameworthy)$ action done by agent, and $(admiration \oplus anger)$ following $(praiseworthy \oplus blameworthy)$ action done by another agent. Emotions are represented via variables of values in 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 gives the intensity of the emotion $j \ni (joy, distress, \text{etc})$. There are three different emotional vectors defined: those triggered by events, the emotional state of the NPC, and which emotions are outwardly (i.e. via animation) by an NPC.

A major aspect of emotions within OSC is how they are affected by personality, for which the paper notes that social psychology research shows personality influences both emotions and corresponding behaviors. While the OSC model does not cover personality influences 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, noting 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. The paper provides some rules for how each of the traits affects the emotional stimulus that an NPC has for an event, and is where the 'Big-5' Personality Traits model comes into play. 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 effect on negative emotions, Neurotics feel negative emotions with a higher intensity than otherwise with no effect on positive emotions, and 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*⊕*sorryFor*) occurs when an agent finds out about a (*desirable*⊕*undesirable*) event that occurred to someone they like; and (*gloating*⊕*resentment*) occurs when an agent finds out about a (*desirable*⊕*undesirable*) event that occurred to someone they dislike.

As to implementation: Impact of Personality on Emotions is computed by use of an Emotional Intensity mechanism as follows. There is first the temporal decay function of which “characterizes temporal evolution of intensity of an NPC’s emotion when no emotional stimulus occurs”. This acts as the main function for which the following three thresholds are further utilized. The Base Intensity Threshold 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 it displays. The Emotion Activation Threshold 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. The 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. Further implementation details for these thresholds are out of scope for OSC and our report, as they encompass affecting how the agent visibly expresses emotions, plans actions, etc.

Social Relations Model

OSC Introduces their social relations model by connecting emotion to 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]. 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, providing recap for some of Fëa’s goals: (1) *A social-psychological model including personality, emotions, and attitudes such NPC behavior is computed based on personality and attitudes;* (2) *A Linguistic style of dialog between virtual characters is determined according to social variables (social*

power/distance) and user's emotions; (3) A Social layer manages communication and biases plan generation and execution in accordance with a social context; (4) Social filters constrain NPCs expression of emotion depending on social context, represented via social power/distance; (5) the NPC's personality, and emotions; and (6) 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 for which OSC only provides limited of support; as it doesn't tread too deeply into realizing interpersonal relations – but 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'). And 'Familiarity' represents how well an NPC knows another NPC/PC based on the type of info and amount shared. Thus the relationship of NPC p to NPC q at time t can be expressed with a 4-tuple as follows $SR(p, q, t) = \langle d^L, d^D, d^F, d^S \rangle$; where the values $\{d^L, d^D, d^F, d^S\}$ are in range $[-1, 1]$ and represent degree of liking, dominance, familiarity, and solidarity respectively, at time t ; where the meanings of these variables correspond directly to their definitions.

OSC further defines social relationships via 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, 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 values. 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.) OSC represents social role values as a lookup table formed by the Cartesian Product Matrix $R \times R$ of the set or roles R whose values are accessible via $SR(p, q, t)$ which would return the rank of the social role of p relative to q at time t . The returned value is a 4-tuple $\langle d^L, d^D, d^F, d^S \rangle$ composed as defined above; except if the roles are NOT complementary, then the tuple is a null vector or has all values set to zero (WLOG). Multiple roles are possible between the same pair of characters, for which OSC takes a simple average.

Event Representation Model

OSC Events are modeled 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' and 'recipient of its effects', which we call "effector" and "affected" respectively. They are represented by a 4-tuple $Event_e = \langle effector_e, action_e, affected_e, time_e, certainty_e \rangle$ composed as defined above in addition to two more variables. The Effector is the entity performing the action. The OSC paper limits this to only NPCs, but we extend it to support groups/factions (e.g. 'Martians invaded Earth'), non-character

living things (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; and note that our effector domain extension is compatible with OSC because such events are similarly handled. Action is the action corresponding to the event. Affected is the entity which absorbs the action. As mentioned, time was added by us, as 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 'calming down' e.g. after getting scared, surprised, etc. We added 'Time' for use with inference (e.g. Temporal Axioms/Beliefs) as a timestamp of either when the event occurred or when it was received.

Degree of Certainty (DC) is a surprisingly value expressing the degree by which agent receiving the event is confident that it actually took place. It is expressed in the range $[0, \dots, 1]$ where 1 implies that the NPC is absolutely sure the event occurred as reported, 0 implies they once thought the event was possible, now knows it did NOT happen, and $(0, \dots, 1)$ implies that the NPC believes event is more or less likely to have happened. To determine the DC of an event received from a fellow NPC (via a dialogue exchange), the paper 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. 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! 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. Regarding implementation, we must likewise restrain from additional details due to the scope of this paper; 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.

Finally, OSC provides two methods for how NPCs ‘hear about’ events: being informed by some other NPC, or inferring an event given clues and/or its KB. 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).

Section 3 – CiF Model

CiF Model Introduction

‘Comme il Faut’ as named by its authors, is the next model, and translates from French to ‘As it Should be’ (hereafter CiF). As opposed to OSC which provides a robust encapsulation for many aspects of an NPC AI, CiF is more focused on social dynamics, thus simplifying aspects such as emotion and personality. However, to recap a theme expressed throughout this report: the ‘beauty’ of the models we’ve chosen manifests in both their ‘overlap’ (i.e. compatibility); and how they ‘fill in the gaps’ of missing desired features among themselves (as to act like modules, and for that matter, even like Knowledge-Source modules within a Blackboard Architecture). Similar to OSC, CiF opens with discussion of its motivations i.e. problem definition and satisfiability requirements. One excellent example describes how social activity encompasses ‘social actors’ which can respond to, update (i.e. change), and display (i.e. exhibit) social state. It makes the excellent point and a major qualifier of consistency in that *“complex social actions are*

best done through dialog... to refer to feelings, relationship states, and history”, qualifying the latter state by adding *“social actions strongly depend on the history of previous social acts.”* Another idea had us draw an interesting comparison with internetworking principles. CiF notes that *“social activity...involves social actors responding to, displaying, and changing social state”*, which led to our analogy that Socially active NPCs are more like nodes in an internetwork versus in a rooted DAG. That is, they are decentralized from each other in nature: thus, need to communicate information between each other about their state, the state of their LANs, and their preferences; while also organizing themselves and effecting cohesion among multiple local and global social hierarchies. This curious idea is further supported by the requirement that *“social action is embedded in a rich social context [such that] actions often have ramifications across multiple social actors”* of which we include multiple social groups (which CiF realizes, if only implicitly). This is likewise similar to internetworking, in that changes in the social state of an agent within its *{family → neighborhood → tribe → nation}* hierarchies may incur a cascading change among all members thereof and beyond one or two ‘hop counts’; just as router state changes would incur among its various *{switch → LAN → IntraDomain → Internetwork}* hierarchies.

Background Models / Analogs

CiF proposes a *“social simulation architecture”* which satisfies such requirements via implementing a realization that supports multi/inter-character social interaction (i.e. exchanges) given KB’s from two domains: via personalities and attitudes i.e. perspectives (similar to how OSC realizes emotional states); as well as via prior dialogues between the interlocutors of the particular dialogue exchange (similar to a priori inference and the Dialogue Agents model discussed in the next section). CiF’s goal is thus realizing the ability for groups of NPCs to ‘engage in rich social interaction’ via ‘emergent social simulation’ that satisfies the above requirements and aforementioned definition of behavioral consistency; and provide some key implementation features to support this goal. The first is that multicharacter social exchanges are represented independent from any character and targeted towards specific characters WRT the

personality traits, social state, and we'll add 'social role' (from OSC) of the initiator NPC for the exchange p and that of the receiving NPC q as relative to the known facts/beliefs p has for q applicable thereof. Basically, this realizes the concept of 'Zeitgeist' in that no interlocutors 'own' the conversation because while they do have volition over what contributions they choose to make: the conversation nonetheless exists not even as the sum of all interlocutors, but outside of them as its own construct.

There are some other features worth discussing. One is that NPCs engage socially via 'soft decision making' (i.e. not Boolean flags nor rigid preconditions) as for such decisions to be made through taking many considerations into account. This is a sufficient condition for the necessity of a 'conscious' agent that is able to both quickly react to social exchanges (without a readily available lookup table) and deliberately plan others (e.g. complex exchanges such as conflict resolution). Also, CiF realizes the full potential of OSC's Degree of Certainty and transmission of events therein. Lastly: previous social exchanges are stored in memory and used both during social exchange performances (i.e. current dialogs can refer to past ones) and to help determine which exchanges an NPC wants to partake in versus rejecting. In addition to supporting the previous feature, this also involves KR/KB aspects and Bayesian Reasoning to evaluate, a priori, whether to partake in or invoke future conversations vis-à-vis predictions of how the other agent would react. Combined: all of these motivations, requirements, and features establish a behaviorally and contextually consistent social simulation whose characters dynamically and constantly evolve their social state via their changing interpersonal relations among a variety of social groups they participate in as well as with events in their world; thus filling in several missing pieces of OSC's personality/social realization.

CiF Architecture Overview

The central KR element of CiF is called a 'Social Exchange' and is composed of a "*collection of dialogue interactions*" structured as to support variable social performances and outcomes relative to the personality traits and social state of the involved interlocutors. They derive from a psychological model

developed by Sociologist Erving Goffman who compared social interaction to theatrical performances wherein characters' personalities were emoted as to change the social state; as well as how the 'setting' and 'cast' for a social interaction has a 'cast of characters' where some "draft others into their performances" as well as both 'stages' and 'props'. We must note that before our discussion needed to be pruned of goal/action planning and decision making: this inclusion of 'props' would have coupled CiF extremely well with Needs-Based AI. This is due to latter models' implementation of a key mechanic called 'Object Advertisement'; where the needs-oriented (and even the emotion-oriented) utility i.e. benefits of an NPC using or interacting with an object is 'advertised out' to all NPC's from all objects (WLOG to efficiency heuristics i.e. spatial partitioning data structures). In any event, the concept of comparing NPC social exchanges to theatre is profound despite our initial criticism, as it focuses social exchanges towards modelling the *"taxonomy of performances organized around what elements of social state they are designed to change or express"*; i.e. a semantics for describing the narrative of a particular conversation between arbitrary agents as for the outcome thereof to vary WRT placing specific agents therein, as seen in activities such as mad-libs and improv theatre (e.g. 'Whose Line Is It Anyway?')

NPC Character Representation

CiF represents the state of a character via 6 elements of various type. The first two are primitives for the character's name and gender. The next three are all vectors encompassing the character's [personality] traits, [social and emotional] status, and 'Prospective Memory' which is analogous to its desires. Lastly, there is a dictionary of 'character-specific phrases' as will be used to 'fill in' certain elements of a social exchange construct as discussed immediately above in the previous sub-section. CiF explains that this state is designed to be lightweight as to carry only a *"small amount of declarative information, i.e. describing who the character is, but not how exactly how they should act"*, with the rationale that *"what makes a character rich and unique is their situation in the social world and history therein, rather than a bunch of character-specific AI behavior."*

Traits represent the personality of an NPC, expressed in CiF via a vector of ‘permanent’ character traits; for which OSC’s is somewhat equivalent (though I loosen the definition for both to be ‘long-term’, even permanent); and are likewise utilized to guide the reasoning/inference process. Statuses represent the state of an NPC, and are expressed as a *“temporary, directional, binary social effects that result from social exchanges”* which encompass the transition of social and emotional states over time. They Capture Transition States in agent’s mood e.g. $\{... (neutral, T^x), (happy, T^{x+1}), ... \}$, spikes in emotion between agents e.g. $status(threatens, a_x, a_y)$, social states of an agent e.g. $status(popular, a_x)$, duration of status before timing out, and timestamp of when the status was effected (as temporal axiom in complement Duration). We note that the use of timestamps is analogous to OSC event. Prospective Memory represents a *“Numeric Vector of Desires i.e. Volitions”* of which ‘drive’ characters to engage in certain social exchanges with certain characters. These are interesting as they tie into both Dialogue Agents as well as OCS WRT a somewhat analogous measure of via ‘Initiator Influence Rules’ (discussed below). Lastly, Character-Specific Phrases populate appropriate ‘slots’ within dialogue templates of Social Exchange constructs, e.g.: $\{ \%greeting\%, \%positiveAdj\%, \%pejorative\%, \%shocked\%, ... \}$

Social State Representation

CiF utilizes 4 different KBs for social state: Social Networks, Relationships, Cultural KB, and the Social Facts KB. We begin with discussing Social Networks, described as *“bidirectional fully connected networks where edge values in range $[0,100]$ measure the feelings between one character and another WRT the direction of the edge”*. There can be many such networks, each encompassing a particular feature (e.g. ‘romance’, ‘friendship’, ‘coolness’, ‘respect’, ‘jealousy’, ‘trustworthiness’) which satisfies support towards realizing a dynamic definition of networks as for new aspects to be added over time while others expire. The OSC equivalent clearly appears to be their ‘social variables’, except they have a fixed i.e. ‘static’ definition at initialization. To clarify the bidirectional aspect: it implies differing opinions between characters such that

$Edge(friendship, x, y) = \{network('friendship', x, y, 75), network('friendship', y, x, 25)\}$ means that x considers y to be their friend more than y considers x to be their friend. It is also similar in terms of effect on social interaction with OSC's social variables, in that they each act as drivers towards building a relationship, but do not necessarily represent such a relationship standalone. For example: that two agents share a mutually high friendship value does not necessarily correlate to everyone else knowing that they're close friends, nor does it mean that they must be or become friends. It simply means they are each inclined to engage in social exchanges that will maintain or grow their friendship. Thus, Social Networks values represent "*private feelings characters have for one each other*" i.e. are not meant to be broadcast to other NPC's; though the 'Relationships' KB discussed below is meant to be 'publicly available' information, and OSC offers further mechanisms that are consistent via its social variable of 'familiarity'.

Relationships encompass the second KR, and realize 'publicly recognized' i.e. 'advertised' social statuses between NPCs. Unlike Social Networks, Relationships are of a binary form, i.e. 'is or is not'. Examples of relationships include 'friends', 'dating', 'enemies', etc. Note that this representation supports the affirmative conjunction of 'friends' and 'enemies' as to represent more complex relationships (i.e. 'frenemies'). Lastly, a powerful aspect of these KRs is that their combination can be used to represent complex relationships. The paper provides a great example of a complex relationship between three characters $\{x, y, z\}$ encompassing the case where $\{x, z\}$ are currently in a dating relationship such that $\{x\}$ has a low romance association with $\{y\}$ yet a high romance association with $\{z\}$; while $\{y\}$ still has a high romance association with $\{x\}$. In other words syntax: $\{x, y\}$ are a couple where $\{x\}$ fell out of love with $\{y\}$ and is now in love with $\{z\}$, yet $\{y\}$ is still attracted to $\{x\}$. When combined with the 'Initiator and Responder Influence Rules', this can lead to a variety of ways in which this situation can evolve: from a love triangle to $\{y\}$ deciding to win back $\{x\}$, and even $\{x, y\}$ amicably deciding to break up s.t. $\{x\}$ is only

a small bit sad. Lastly, this is how the paper encodes this scenario: $\{relationship(dating, x, y) \wedge net(romance, x, y, 20) \wedge net(romance, y, x, 95) \wedge net(romance, x, z, 80)\}$.

The Cultural KB (hereafter CKB) composes the third KR, providing a *“sociologically rich representation of props [ergo] variety of topics to bond over and squabble about”*. This includes objects, concepts, genres, and we also add events in the world. Basically, it helps implement topics for the agents to talk about and evolve relationships over, and from our inclusion of events: is also an emergent, implicitly procedurally generated source of discussion. Standalone, we realize an improvement to the pre-designed, static dialogue trees implemented by present day AAA open-world RPGs. To augment this idea into an even more powerful but just as simple realization: we propose assigning priorities and descriptive words to CKB entries in addition to the other contents described immediately below such that, for example, the students of a simulated High School will gossip about “recent news that is ‘cool’ or ‘hot’ versus older news considered ‘boring’ and ‘depressing’”. We make the usual comparison in noting that equivalent of the CKB for OSC model could be Praise/Action outputs based on personality traits. This is especially as the opinion for each CKB object can be universally agreed-upon just as with Action-Effects in OSC, or subjective to an individual NPC’s personal opinion/preference: as with Praiseworthiness in OSC. CKB entries are represented as connections from an agent to one or more objects in CKB via unidirectional edges encompassing negatable verbs (expressed disjointly in the paper) e.g. ‘likes/hates’, ‘wants/has’, etc. The reason I mention how the CiF model has them disjoint is that its CKB relates to OSC’s Actions/Praise; except the latter is more robust i.e. of range [-100,100] ergo does not partition a verb from its negation.

The Social Facts KB (hereafter SFKB) records every social exchange made between characters and every ‘trigger rule’ (event?) which causes social state change, as to *“keep track of social history of story world so that it can be queried for socially relevant information”*. This, in turn, can be used to drive the types of

social exchanges characters choose to participate in and how the resulting exchange is driven WRT previous exchanges. That is: it realizes bringing up prior conversations characters had with each other and even with other characters. The paper makes a very important observation in that games don't record much of the interaction history if any for the purpose of decision making or to cite in future dialogue exchanges; for which we add that the only mechanisms in our experience that come close are how Bethesda's Open-World titles, conversely, seem to only realize this via setting trigger flags. SFKB Entries are composed of *"details of who was involved, time of exchange, items from CKB that were mentioned, natural language generation template for turning entry into text for use in dialogue, and social exchange labels (i.e. 'mean', 'funny', 'nice to') that can be used for querying.* A final powerful idea of SFKB entries are their 'compounding effect', wherein SFKB labels can be used to realize 'remembering history of past exchanges' in terms of one or more themes e.g. 'when were they nice to me? When did they act stupid?' As the paper puts it: *"SFKB supports a compounding effect of history, where the characters refer more and more to past events that have happened, with the past events affecting decision making"*

Social Exchange Representation

As mentioned in the Architecture Overview subsection, social exchanges are at the core of CiF because they effect transitions between social states via social state. We define the parties of a Social Exchange with different words than CiF that we find more intuitive, they are the Initiator (I), the Responder (R), or an optional 3rd agent (O). The composition for a Social Exchange includes its name, intent, precondition (optional), initiator/responder influence rules, and instantiations. The intent is the purpose that the initiator has for initiating the exchange i.e. changing a social network value or relationship status. The preconditions are optional conditions that must be true for the social exchange to be applicable in the current social environment; and tie into analogous concepts from the next section on Dialogue Agents.

Instantiations are composed of effects and natural language generation templates, as divided into 'accept' and 'reject' instantiations; s.t. each is mutually exclusive to how a social exchange can play out. Associated with instantiations are conditions that are tested to see if the instantiation is valid in the current context. Every exchange has a generic accept and reject instantiation that places no conditions upon it. More specialized instantiations have additional conditions, and play out exchange in specialized ways mand with more side effects. The instantiation performance consists of dialog represented using natural language generation templates to be spoken by the characters during the exchange (stuff like animation / emote tags can be tagged as well). The paper uses the example of character 'x' asking character 'y' out on a date, for which 'y' is not compelled to accept and may instead both reject the request for any number of reasons as well as 'penalize' the action by decrementing certain social network scores towards 'x'. Furthermore, as with 'Compounding Effect of SFKB Entries': there can be a cascade effect of this event unto other social state; for example: feeling embarrassment unto the rest of some group, and/or other NPCs 'hearing' about the event and feeling pity towards 'x' (which ties into OSC in terms of 'Degree of Certainty' and its quasi-model of empathy vis-a-vis 'feels concerned [for]'. Similar effects are possible if the request is successful (e.g. character 'p' has a crush on 'y' and feels angry/jealous that 'y' accepted dating 'x').

Influence Rules are composed of predicate arguments such that there are sets unique to specific characters, and sets that are universal for all. A possible analog of these rules could be the fundamental nature of beliefs -vs- axioms; where beliefs may be unique to an individual where axioms are considered to be universally understood ideas. These rules are used to determine volition as well as whether to accept or reject entrance into a social exchange (i.e. encompass decision-making/action-taking system). They further seek to combine advantages of utility methods and Boolean predicate representations (pretty much as discussed in class). This 'Predicate Calculus' warrants a note or two: it's great at representing complex conjunctions of states, which we know is hard to express through combining algebraic functions.

By having predicate calculus expressions on the left-hand side of rules, but having them add weights to a sum on the right-hand side, the rules work together to compute a complex utility surface, combining the benefits of numeric and logic-based decision making. There are two types of influence rules as to compose a second dimension alongside unique-vs-universal rules. The first kind are initiator rules, which determine the volition (i.e. desire) of a character to initiate social exchange with another character; i.e. *“social considerations applicable to initiator of this social exchange.”* The second kind are responder influence rules, which determine whether a responder accepts or rejects social exchange i.e. *“social considerations applicable to the responder of this social exchange.”* The paper notes that responder rules are similar to initiator rules, except they are used to score how responders ‘feel’ about the exchange they’re involved with. This engenders a process similar to desire formation, the responder gets to determine how they feel about the exchange. If the responder score is too low, the responder will reject the exchange, resulting in a different (and often opposite) social effect than social exchange intent. In addition to exchange-specific responder influence rules, microtheories are also used for computing the responder score.

Section 4 – Dialogue Agents Model

Dialogue Agents Introduction

We follow our synopses of the robust OSC and CiF models with two additional ones of which within the scope of this report are only partially utilized as to extract smaller components thereof for particular use with personality and social simulation; but note that for a full realization of the Fëa composite model: each would be more substantially utilized. The first of these is the Dialogue Agents model which is not necessarily a model as much as a summary paper identifying design patterns for dialogue within video games compared with corresponding case studies. We utilized it WRT to the scope of this paper as to provide a bit more context into the nature of communication between both NPCs with each other, as well as with PCs (i.e. humans). It foremost defines conversation dialogues between agents in video game /

virtual simulation worlds as a “means for agents to exchange and coordinate information, build social cohesion, and achieve mutual understanding”, for which CiF provided similar observations.

Dialogue Types and Elements

Two main types of dialogue are defined, as well as three types of ‘elements’ which compose them (of which we’ve slightly adapted for use with Fëa). All three elements are familiar from the discussion so far: PCs which are human players, NPCs which are autonomous agents, and OCE’s which are ‘Observed / Communicated Events’. We admit that OCE’s bend the convention a bit but are important as to help unify the representation of information and events between this model and OSC, CiF, and Needs-Based as also presented in this paper / part of our conceptual composite model. OCE’s represent either an observation that has been made by some character (e.g. ‘NPC x sees arriving UFO’) or an event that has been communicated to a character from something other than another character (e.g. ‘NPC y watches a news report about the arrival of a UFO’) For the purposes of our brief discussion on Dialogue-Agents given the limited scope: OCEs are only mentioned without further immediate consideration. As to the dialogue types, they follow from the elements. [PC-NPC] encompasses a human having a dialogue exchange with an NPC. The paper partitions human players (i.e. P) from the player-characters a.k.a. ‘avatars’ whom players control within the game world (i.e. PC), but this is unnecessary for our purposes, thus we resolve both accordingly into the single type. The purpose of such dialogues is classically to provide the player with exposition, quests, and other information, while having a peripheral purpose of ‘filling empty space’ by providing some basic social exchanges from pre-baked dialogue trees. The other dialogue type is [NPC-NPC], which encompasses an NPC having a dialogue exchange with another NPC. The purpose of these dialogues are classically to comment on the states and events of the game world, for which most implementations even to present realize in a very limited sense (e.g. via locking and/or unlocking pre-scripted dialogue trees based on activated triggers specifying the occurrence of world events). CiF and OSC, however, propose a vast expansion of the effects of dialogues between both PC-NPC and NPC-NPC.

This expansion includes dynamic social context/state changes (i.e. conversation goes badly if player gives NPC bad news), types of dialogue available WRT robust personality simulation and emotional state (i.e. not just limited to unlocking new dialogue if 'charisma' stat is >50, can realize WRT "PC is confident, has a friendly demeanor, and is dominant over the NPC; NPC is agreeable, somewhat likes the PC, has good solidarity with PC, and does not feel threatened by PC – such will unlock the ability to ask several questions for which the NPC will not otherwise be willing to answer"

Agent-Based Dialog / Turn-Taking

There are two important and somewhat coupled topics within this paper, for which the first provides us with a definition for Agent-Based Dialog Systems given the aforementioned introduction. The second idea is 'turn taking', which encompasses how 'moves' (ply's?) in a conversation are organized and distributed. Ergo to begin: the paper establishes an 'Agent-Based Dialogue System' as one which satisfies a principal requirement that follows from our main theme/requirement of 'Behavioral and Contextual Consistency'. Specifically, it states that (paraphrased due to brevity) '[realistic] conversation requires agents with the ability to re-plan towards achieving existing goals xor defining new ones whenever its priorities change - especially for dynamic simulation worlds wherein preconditions and/or circumstances required to execute the current plan or goals may frequently change". We mostly bring this up to, once again, connect the goal- planning / action-taking / decision-making behavioral systems not explicitly discussed within this report but whose shadows still affect our full realization of a consistent 'NPC Framework' as is our goal for the full Fëa model; given that those systems are affected by and affect personality / emotion / attitude / social simulation models that the partial version of Fëa modelled in this paper does involve.

There are two kinds of turn taking discussed in the paper. The first is 'Incremental Dialogue Processing' of which encompasses human-to-human conversation, wherein interrupts "are possible [because] the hearer can process and react to each contribution while it's being produced." Chunk-Based Dialogue

Processing involved a more restricted alternative which only allows communication via "utterances or phrases" (a.k.a. 'blurbs of text') transmitted as a whole; which has been and remains the convention for many open-world RPG games for which we aim to drastically improve via models like Fëa. In either case, we note that the spaces between moves is analogous to state transitions, for which we discuss further WRT implications with OSC/CiF in 'Affective Dialogue' below. However, there is a middle ground between Incremental and Chunk-Based dialogue processing, especially vis-à-vis NPC-NPC dialogue exchanges, in the form of Interruptible Actions. Such actions allow an NPC to interrupt an ongoing conversation (perhaps even those of which it is not [yet] involved i.e. to reject/refuse participation in a conversation). Of this, we first note the clear connection between the mechanism of NPC's rejecting conversations given their attitude and personality via both the CiF and OSC models, their emotions given the OSC and upcoming Needs-Based AI models, and their memories of prior exchanges given the CiF model. Furthermore, we note that via structuring the interruption as an event (i.e. by definition via the OSC model), we could further implement the ability for the interlocutors to express objection to either a rejection for conversation by a receiving NPC (e.g. as to 'bark' back a response such as "Hey get back here! You're going to listen to what I have to say!") xor an interruption between two interlocutors by a 3rd party (e.g. the classic "How rude! This conversation is between 'A' and 'B', so why don't you 'C' yourself out of it!")

Affective Communication

Affective Communication is a crucial enhancement to dialogue methods of which helps satisfy consistency. It involves influencing the NPC's disposition towards the PC (or other NPCs for an NPC-NPC dialogue) via emotion-based responses (e.g. Admire / Intimidate / Taunt / Bribe). There is already precedent with affective communication as can be seen with many of the open-world RPG games such as Bethesda's 'Fallout' and 'Elder Scrolls' serials; despite remaining relatively limited/simplistic by largely involving pre-scripted dialogue and event sequences versus the kinds of emergent, procedural realizations sought by Fëa and its module components. However, we note from this precedent clear examples for how

such dialogue options can affect game state by doing everything from setting a character hostile to activating some event trigger in the game world; and especially: changing the flow of not only future conversations but the current conversation (i.e. via either locking xor unlocking dialogue choices). With a fully procedural and emergent realization, these effects stand to be rapidly enhance, encompassing two major implications for such a realization. The implications on NPC Personality involve a much more 'familiarity-driven' approach for how a character can persuade another character to become friends, provide information, do or don't do a certain task, etc. This extends to NPC-PC relations as well; as a 'crazy' possibility for robust enough NPC intelligences is to actually turn the tables around and manipulate the actions and thoughts of human players! This would encompass an analog phenomenon similar to what we call the 'Iteration 2 Turing Test' as inspired by the 2014 film 'Ex Machina', wherein the goal is for an artificial agent to successfully deceive a human who is clearly aware that they are engaging with an AI.

Section 5 – Needs-Based AI Model

Needs-Based AI Introduction

The Needs-Based AI Model is effectively a derivative of the model used with the classic video game 'The Sims'; as the author of this model helped develop the game and 'cousin' expansions of its model for other titles since. Although we will mostly be referencing the 'Physical Needs' component, we note that it's only a piece of a far more powerful model/system which further implements action selection/performance, realizing complex 'action chains', and some degree of goal planning. We also refer to a presentation given on this model alongside the greater AI implementation of 'The Sims' of which discusses some of the social dynamics/relations mechanisms. We also need to discuss it because physical needs to affect human emotions and vice versa. If we have implemented the simulation of physical needs into our NPCs and desire a realistic and accurate realization of their personality and emotions: needs-based AI is essential!

Needs and Types thereof

OSC and CiF utilize several major psychological models, of which the 'Big 5 Personality Traits' and OCC are well-known. Another well-known model that some may comment has been left out of our discussion so far is 'Maslow's Hierarchy of Needs' – but now it gets its turn. The Maslow model is a “theory of human behavior based on increasingly important psychological needs” from which 'The Sims' modeled its “biological and emotional drivers” for NPC agents. Needs-Based AI is thus a model by which action selection (typically immediate i.e. short term) is based on fulfilling a set of needs whose units represent the 'motivation' of satisfying each such need (towards some homeostasis 'balance' between the needs WLOG). Needs are represented as an array of numeric values of range [0, 100] which decay over time; and increase/decrease based on actions and events. They have the semantics of “lower is worse and more urgent”, so that hunger=30 means “I’m pretty hungry,” while hunger=90 means “I’m satiated.” Need values should decay over time, to simulate unattended needs getting increasingly worse and more urgent. Performing an appropriate action then refills the need, raising it back to a higher value (this is a connection from actions to need satisfaction, ergo likewise resolution of impact of needs on mood, emotions, etc.) For example, we simulate the agent getting hungry if they don’t eat, by decaying the hunger value over time. Performing the “eat” action would then refill it, causing it to become less urgent (for a while).

There are 8 types of needs separated into 2 groups: physical and mental. Physical needs include hunger (remedied by eating), hygiene (remedied by cleaning), bladder (self-evident), and comfort (remedied by sitting / laying down). Mental needs include energy (remedied via Sleep Cycles of which itself is a VERY interesting and realistic mechanism but sadly out of scope of this paper), social (remedied by interaction with other NPCs/PCs), fun (remedied by entertainment / recreation, and environment a.k.a. room (remedied by architecture/aesthetic). Regarding environment: the details are left vague but we believe this simulates things like becoming depressed if one lives indoors all day without any exposure to sunlight,

lives in a derelict ruin versus a luxury apartment, etc. Regarding 'energy': we are compelled to express a contention for which we also offer a resolution thereof. To address this in reverse order: we believe that 'energy' could be partitioned into 'physical' and 'mental' such that each certainly affects the other; while each would be affected by only their corresponding needs. That is, low hunger would affect the decay of 'physical energy' which would then affect 'mental energy' such that low hunger affects mental energy - but only by proxy of affecting its respective energy need. To provide intuition to this idea: being tired after a 5-mile jog does not necessarily imply a corresponding fatigue for solving math problems; nor does completing a math exam imply a corresponding fatigue for the 5 mile jog. While each task certainly affects the energy available for other - we do this only through one energy need affecting the other. That is, taking the exam reduces mental energy, which then reduces physical energy for jogging - and vice versa.

Implementation within AI system

We conclude our synopsis of the Needs-Based AI model with some implementation notes, noting the obvious in that these notes are sparse. This is largely due to the fact that as with Dialogue Agents, only some components of this model apply within scope of this project. For a full realization of Fëa and all of its components: a substantially greater discussion of implementation for this model would be needed. That said, we first note that the means by which needs values are turned into a 'mood score' again utilizes principles in behavioral mathematics. Specifically: the sum of all physical and mental need values is taken such that each is weighted by respective hyperparameters which tune their influence upon the computed mood score not unlike a simple classifier network. Going from the Will Wright pseudocode as we implemented in a basic demonstration of the 'Sims-style Needs Model' back in 2019: the resulting subtotal score is then factored against a fraction (technically – is divided by a whole number) such that this number is also effectively a hyperparameter (in Wright's case – 17) from which a final mood value is produced. The next question is what the greater Fëa system does with this mood value; for which we have an idea. One option is to have individual needs values both explicitly affect their corresponding emotions and

personality types (i.e. low social need value and extrovert means joy decreases); while also using the mood value as a standalone emotional component which speaks for the character's "mood given needs" whereas the OSC value produces "mood given emotions". We could then either take a weighted average of both, or experiment with a number of such calculations WLOG.

Two final points involve the topics removed from scope of this report; but of which remain applicable due to their prospective impact on emotions given needs if/when a full realization is modelled/realized. The first regards the impact of needs on actions and behaviors thereof. We observe that 'The Sims' model uses simple utility calculus on the rewards of all possible actions s.t. the action chosen is defined as the "random [choice] among the top 4 that cause the most happiness [i.e. utility]". In response to this, we instead suggest a reactive intensity/priority model could be modeled implemented analogous to how OSC calculates emotions, wherein more intense higher priorities (i.e. really hungry and hunger is high priority need, need to piss really bad but enjoying television show) are chosen consequent, without going into any more detail. Finally, we make a small remark about the impact of Personality attitudes on Objects (standalone and WRT advertisement concept): in that this could allow for realizations such as "being unhappy because living in dirty house" but happiness would increase if house was cleaner, etc.

Section 6 - Conclusion

This report discussed the simulation of Personality, Emotions, Social Relations, and the Physical/Mental needs of NPCs by providing a synopsis on four models: OSC, CiF, Dialogue Agents, and Needs-Based AI; and connecting where these models overlap and diverge beneficially towards the conceptualization of a composite NPC AI architecture containing elements from each, which we named Fëa. We also discussed how these models would utilize and be utilized by a greater AI system with elements such as Knowledge Bases, Inference Reasoning, Goal Planning, and Action Tasking; while noting that our report needed to

largely omit further details as to keep within reasonable scale and scope. Indeed, it turned out that some of our models did the same thing. In the case of OSC, they note: "the benefits of our enriched modeling will indeed appear more clearly when it [is integrated with] a full model with action selection and behavior control", of which we can completely relate! We discussed the importance of consistency and the use of established models from social and cognitive psychology to establish an accurate simulation in terms of both context and behavior. We discussed several major applications for NPCs of such intelligence, ranging from industrial training and pedagogy, to robotics and virtual assistants, and of course: entertainment. In conclusion: we feel that this first significant expedition into the realization of Fëa has achieved our primary goals as discussed in the introduction; while providing ideas and a foundation from which to iterate upon for possible future work, including an incremental implementation of the composite model idea.

Section 7 - Bibliography

M. Ochs, N. Sabouret and V. Corruble, "Simulation of the Dynamics of Nonplayer Characters' Emotions and Social Relations in Games." *IEEE Transactions on Computational Intelligence and AI in Games*, Vol. 1, No. 4, pp. 281-297, Dec. 2009.

M. Mateas, and J. McCoy, "An Architecture for Character-Rich Social Simulation." *Game AI Pro*. Volume 1, edited by Steve Rabin, CRC Press, 2014, pp. 515-529

R. Zubek, "Needs-Based AI." *Game Programming Gems 8*, edited by Adam Lake, Cengage Learning, 2011, pp. 302-312.

J. Brusk, S. Björk, "Gameplay Design Patterns for Game Dialogues", Proceedings of the 2009 DIGRA International Conference, London UK, September 1–4, 2009.

Mark, Dave. *Behavioral Mathematics for Game AI*, Cengage Learning, 2009.