## **CiF Model Introduction**

The second model we researched for integration into Fëa’s composite model is ‘Comme il Faut’ as named by its authors, which translates from French for ‘As it Should be’ (and for which we hereafter shorten to the CiF Model). Versus OSC which provides a self-evidently robust encapsulation for many aspects of the greater personality/social realization we strive for with Fëa, CiF is more focused on social dynamics, leaving aspects such as emotion and personality more streamlined. However, to recap the theme expressed throughout this report: the ‘beauty’ of the models we’ve chosen manifests in both their ‘overlap’ (as to be compatible with one another in terms of KR’s and other mechanisms); 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, the CiF paper opens with discussion of its motivations i.e. problem definition and requirements for a satisfiable solution method. An excellent example describes how social activity encompasses ‘social actors’ which can respond to, update (i.e. change), and display (i.e. exhibit) their 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.”

A further motivation made us draw an interesting connection with the fundamentals of internetworking, especially in terms of packet switching. CiF notes that “social activity...involves social actors responding to, displaying, and changing social state”, for which we draw the comparison that Socially active NPCs are more like nodes in an internetwork versus in a rooted DAG: in that 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” for which we include multiple social groups as well (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 of those domains beyond a small constant hop count as router state changes would incur among its {switch > LAN > Intra-Domain > Internetwork} hierarchies.

## **Background Models / Connections to OSC, Dialogue Agents**

Thus, 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 and that of the receiving NPC as relative to the known facts/beliefs has for 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. in the case of complex exchanges such as conflict resolution or oppositely: notifying an adversary about an act of revenge including the motivations of why whether or not such were asked for, as discussed below). Relative to the Fëa: this also involves decision making / reasoning given NPCs attitudes, emotions, and general mood; and can be very powerful when merged with the other components of our composite model. The other two features follow from the aforementioned goals of the CiF Architecture. One recaps that ‘social interactions don’t just cause a single change in social state, but have cascading consequences across multiple characters.’ As discussed above, this realizes the propagation of the effects of a social exchange throughout the population interested in it or who are informed about it. Furthermore: it also realizes the full potential of ‘Degree of Certainty’ and transmission of events within the OSC model. The other is that previous social interactions 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 implementation features thereof establish both 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 it, as well as with events in their world; thus also realizing several missing pieces of OSC’s personality/social realization. We will now provide a synopsis of the CiF Architecture as to demonstrate some of its mechanisms while connecting them to OSC and the other partitions of the Fëa composite.

## **CiF Architecture Overview**

The central KR element of CiF is called a ‘Social Exchange’ and is composed of a “collection of [primarily dialogue] interaction” 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. The paper 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”, and that their rationale for the information that is kept is thus 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.” Furthermore, we make the usual reminder to note that these definitions are compatible with the main types of general KR such as SOL, as could be presented in scripts via JSON, XML, etc. WLOG)

Traits represent the personality of an NPC, expressed in CiF via a vector of ‘permanent’ character traits; for which OSC has near exact equivalent (though I loosen the definition for both to be ‘long-term’, even permanent). Also, as with OSC: they’re effectively utilized to guide the reasoning/inference process i.e. such is their “meaning”. Statuses represent the state of an NPC, and are expressed as a “temporary, directional, binary social effects that result from social exchanges” of which encompass the transition of social and emotional states over time. They Capture Transition States in agent’s mood e.g. statust-1(neutral, ax)↦statust-1(cheerful, ax), Emotion Spikes between agents e.g. status(threatens, ax, ay), the social states of an agent e.g. status(popular,ax), the duration of how long status exists before timing out, and the timestamp of when the status was effected (as temporal axiom and complement to Duration). We note that the use of timestamps is analogous to OSC event quadruplets. Prospective Memory represent 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 the OCS model WRT a somewhat analogous measure of via ‘Initiator Influence Rules’ (discussed below). Lastly, Character-Specific Phrases populate the appropriate ‘slots’ within dialogue templates of Social Exchange constructs as described in the previous subsection. As mentioned above, they basically work with Social Exchanges as one would in a game of ‘Ad-Libs’ WRT a custom vocab base; e.g. as seen with text ‘blurbs’ in Sid Meiers’ Alpha Centauri of which CiF has similar syntax e.g. {%greeting%,%positiveAdj%,%pejorative%,%shocked%,etc.}

## **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’, etc.) 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: we mean that it implies/realizes 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. Further, similar in terms of effect on social interaction to 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 either the rest of the ‘public’ recognizing them as friends; nor does it necessarily imply that they automatically become friends. It simply means that they will be inclined to engage in social exchanges which will foster the actual relationship as will be recognized by them mutually and the public; and just the same for all other social network types e.g. ‘romance’.

An important note regarding the Social Networks values is that within CiF: they encompass the “private feelings characters have for one another” i.e. are not meant to be openly ‘advertised’ 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}. IOW: {x,y} are a couple where {x} fell out of love with {y} but has the hots for {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 {y}, and even {x,y} amicably deciding to break up s.t. {x} is only a small bit sad (hello more robust personality mappings of OSC system to complement!) BTW: the KR for this scenario is as follows: {relationship(dating, x, y) Λ network(romance, x, y, 20) Λ network(romance, y, x, 95) Λ network(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, ideas, concepts, genres, and we also add events in the world, including relationship changes between other characters. Basically: the CKB helps implement stuff 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; and is standalone a much more consistent than the pre-designed, static dialogue trees implemented by major open-world RPGs even to present. To augment this idea into an even more powerful but just as simple realization within Fëa: 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 object in the CKB 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’, ‘envies/pities’, ‘aroused/disgusted’, etc. The reason I mention how the CiF model has them disjoint is that the CKB therein relates in a sense to OSC’s Actions/Praiseworthiness; except the latter is more robust i.e. of range [-100,100] and thus does not partition a verb from its negation, e.g. {‘likes’, -25}(x, ‘coffee’) ≈ (‘coffee’, (x, hates)=[True] CiF’s CKB is more oriented towards steering conversations; e.g. CKB queries have up to 4 parts such that full query CKB(item,(x, likes), (y, dislikes), lame) “will match an item that x likes, y dislikes, and is universally regarded as lame, which could perhaps contribute to y’s volition to poke fun at x”

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 seem to only realize this via setting pre-set trigger flags. This method allows NPCs to ‘remember’ previous dialogue options selected by the PC only because a flag value was set as to, for example, cause the subtree of some dialogue tree to be unlocked, consequently revealing the corresponding dialogue within. 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”. An example based on the one provided in the paper is expressed as: (SocialGameContext exchangeName = “Bully” initiator = “Edward” responder = “Chloe” initiatorScore = “15” responderScore = “10” time = “5” effectID = “10” other = “”(SFKBLabel type = “mean”)) . This encompasses the fact that ‘Edward made fun of Chloe’s SAT score’; such that it “records the exchange’s name, characters involved, initiator/responder influence rule scores, timestamp, effect ID indicating which specific instantiation of Bully happened (i.e. making fun of SAT scores), and label indicating this exchange was ‘mean’. 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?” This can then be used to implement stuff like “remember all those times when you were mean to me?...” 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” Paper Example - Revenge: Chloe was made fun of by Edward because of her SAT score, and later initiates a ‘DiscloseRevenge’ exchange with Edward of which queries for past mean things he did or said to her: [SFKBLabel(mean, responder, initiator,0) window(10)], in this case looking through past 10 social exchanges. The matched entry is then turned into text, resulting in: “You know when you made fun of my SAT score?”, which the initiator (Chloe) says before revealing how she got revenge.

## **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 e.g. ‘changeSocialNetValue’, ‘changeRelationshipStatus’, etc. 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 on the instantiation. More specialized instantiations have additional conditions, and play out the exchange in more specialized ways, in addition to having more effects. If multiple instantiations have conditions which evaluate as true, the most salient is chosen, with saliency being computed as a weighted sum of the number of true predicates in each condition (the weight associated with a predicate type indicates how important that predicate is for determining the specificity of a social context). The instantiation performance consists of lines of dialog, represented using natural language generation templates, to be spoken by the characters during the exchange. (and stuff like animation / emote tags / calls can be tagged as well) OKAY SO this is IMPORTANT IDEA because it’s what allows for specific reminiscing as shown in example bottom PDF 13 top PDF 14. The details of a specific performance include the NLG dialog templates and character animations. Instantiations have a condition that must be true for it to be performed and a specification of how the performance changes the social world. Lastly, as discussed above: exchanges may be rejected by the responder, which typically causes an opposite state change. The paper uses the example of character ‘x’ asking character ‘y’ out on a date, for which ‘y’ is not automatically compelled to accept but 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 embarassment 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 require a bit of explanation. They are composed of predicate arguments such that there exists sets unique to specific characters, and one set that is universal among all characters (known as ‘Microtheories’ and further discussed below). 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 and as well as whether to accept or reject entrance into a social exchange (i.e. encompass decision-making/action-taking system); and are of the following form: <condition> → <increment/decrement volition for an intent>. 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.

Speaking of microtheories: their purpose is to facilitate knowledge reuse i.e. write single set of rules that can be used for many social exchanges, and can realize even more complex social dynamics such as self-destructive tendencies and ‘behaving oneself for [reasons]’ i.e. avoiding potentially adverse behaviors. They are composed of a definition (as a condition expressed in predicate form) plus a set of influence rules, and the paper discusses implementing a microtheory library consisting of large repository of rules split between dozens of microtheories. To wrap up on microtheories: Only those whose definitions evaluate to true in current context considered when calculating volitions; as for their rule set to then provide a general representation of social “common sense” associated with the condition. Speaking of volitions: they are formed by ‘summing volitions based on satisfied initiator influence rules for every potential responder, producing a vector of numbers corresponding to the character’s desire to engage in each of the social exchanges with each of the characters’. A variety of policies can then be used to decide what action to take, such as weighted random selection from the top N for autonomous action, or placing the top N in a menu for the player to select from as was done in Prom Week. The volition calculation process can be made efficient by limiting set of potential responders (for example to just nearby characters or characters present in a scene) for which volitions are calculated. Lastly Trigger Rules are effects (and associated conditions) that are shared across all social exchanges. In addition to the social state change caused by the social exchange itself, additional changes may be caused by trigger rules. The trigger rules capture the cascading consequences of social exchanges, as well as state changes crossing multiple social exchanges. They are similar to that chart in the OSC paper showing output of tree formed by SOMA quadruplet.