2017 ICCRTS-KSCO

Title of Paper:

A Framework for Modelling the Effect of Emotion on Uncritical Reasoning

ICCRTS Topics:

Cognitive and Socio-technical Challenges
Interoperability, Integration and Security

KSCO Topics:

Coalition and Team Information Sharing
Coordination and Collaboration
Cultural Influences

Name of Authors:

David Mott
IBM UK
Southampton, UK

Troy D. Kelley
Human Research and Engineering Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD, USA

Cheryl Giammanco
US Army Research Laboratory
Aberdeen Proving Ground, MD, USA

Soheil Eshghi
Department of Electrical Engineering and Institute for Network Science,
Yale University, CT, USA

Yunfeng Zhang
Human Agent Collaboration Group
IBM T. J. Watson Research Center, YorkTown Heights, NY, USA

Abstract

We describe research on understanding group mutability in the behaviour of external groups, and how interventions by coalition forces may affect the behaviour in terms of controlling hostile groups and encouraging friendly groups. We explore how emotion may influence the behaviour of individuals by affecting the type of reasoning that they undertake, encouraging "uncritical" rather than "critical" thinking. We describe a computational framework holding a cognitive model of an individual operating within a group context, inspired by theories from social science. Individuals relate to in-groups and out-groups and have beliefs that are associated with emotions. Cognitive Appraisal Theory is used to evaluate incoming memes "pronounced" by external speakers, appraising the effects of the memes on an individual's self-esteem taking account of their group relationships as indicated by social identity theory, and leading to an emotion in the individual. Appraisal is followed by a process of coping that seeks to handle the effects by either performing problem-focussed (critical) or emotion-focussed (uncritical) thinking, according to the current emotional state of the individual. This model is implemented within a Cognitive Architecture (Soar) as a set of reasoning processes that handle beliefs and emotion. The model is integrated into a multi-agent simulation tool (Repast Simphony) allowing the simulation of populations of individuals interacting and spreading rumours, or memes, together with interventions. We describe how this framework could be used to construct experiments to explore how different situations lead to group mutability and behaviour, together with the effects of interventions by coalition forces.

1 INTRODUCTION

One of the aims of the Distributed Analytics and Information Science International Technology Alliance (DAIS ITA) programme [DAIS-ITA 2016] is to understand the mutability of groups to support the achievement of desirable outcomes by the more accurate prediction of group behaviour and the design of effective intervention strategies. This paper outlines research to support the modelling of group mutability and factors causing behaviour, by exploring how emotion can affect the reasoning capabilities of group members. We seek to distinguish between "critical thinking" [Fisher 2011] where logical principles are employed, and "uncritical thinking", where these principles are not necessarily applied. Uncritical reasoning seems to underlie a range of undesirable phenomena such as conspiracy theories, demonisation of individuals and organisations, the refusal to consider expert opinion and the spreading of fear and false information, which in turn has the potential to adversely affect group behaviour. We explore the hypothesis that emotion can have a significant effect on cognition and on the use of uncritical thinking in particular; therefore the understanding of the effects of emotion on cognition is useful in understanding the mutability of group behaviour and how it might be influenced

In our research we are developing computational models based upon cognitive architectures and social theories, that simulate the effects of emotion on cognition in the context of group behaviour and thereby have the potential to provide characterizations of group mutability and the effects of intervention strategies. This work will also support other DAIS ITA research, including the development of a high level model of group behaviour and the development of a meta-heuristic framework for describing different modelling strategies.

The scope of this paper is to describe our computational framework for modelling emotion, cognition and group behaviour, with a focus on critical and uncritical reasoning. We provide an example of a basic experiment using this framework, and discuss how future experiments could be performed, but at this stage we do not provide detailed, scientifically validated, experimental results in scenarios involving complex social phenomena. The paper is structured in nine sections: section 2 reviews previous work in this area; section 3 describes a scenario used to focus the research; section 4 describes some basic concepts and associated social science theories used to develop the cognitive model; section 5 defines the cognitive framework and model in detail; section 6 shows how this framework integrates with a multi-agent simulation; section 7 describes a basic experiment using the framework; section 8 discusses extending the underlying cognitive architecture with emotion-specific modules; and section 9 draws conclusions and outlines future work.

2 PREVIOUS WORK

Previous work in computational models of emotions fall into four main categories: appraisal, dimensional, anatomical and rational [Marsella et al 2010]. The appraisal approach [Smith & Lazarus 1990] suggests that emotions arise from the cognitive agent's continuous assessments of the environment and the relationship between the environment and internal beliefs, behaviours, and concepts. The dimensional approach postulates that emotions are represented in a multi-dimensional space and vary along a continuum within the entire emotional space [Russell 2003]. The anatomical approach focuses on the neurological underpinnings of emotions, attempting to reconstruct neurological correlates and representations that are influenced by brain anatomy [Panskepp 1998]. The rational approach views emotions as another rational mechanism that aids cognition and decision making [Anderson & Lebiere 2003].

We aim for a cognitive approach, with an emphasis on how different forms of reasoning (uncritical and critical) may be influenced by emotion, yet accepting that there may be a logic to "emotional" reasoning. To this end, a combination of the appraisal, dimensional and rational approaches is being followed, rather than consideration of the neurological aspects of emotion.

As described below, our work combines a cognitive architecture (Soar), cognitive appraisal theory, and theories about social behaviour in a group context, with a view to understanding the effects of emotion on critical and uncritical reasoning. Here we survey previous research in the light of these topics. A number of researchers have integrated emotions and cognition within a cognitive architecture, with the dominant approach being based upon Cognitive Appraisal Theory [e.g. Smith & Lazarus 1990], where a situation is appraised using cognitive processes, leading to emotions, and is followed by coping strategies to deal with these emotions.

Much of this work is based upon the Soar cognitive architecture [Lehman et al 2006] to provide the computational framework in which to represent the appraisal process and its effects on reasoning. [Marinier et al 2009] describe an integration between cognitive appraisal theory and Soar, using appraisals such as goal relevance and suddenness to engender emotions and control the construction of actions in a game simulation. However the reasoning involved is aimed at a correct execution of the task; there is no attempt to model "uncritical reasoning" where emotion may "derail" the reasoning. In addition, the appraisals are defined by factors relevant to the game simulation, rather than by social theories about human behaviour in a group context. [Marsella & Gratch 2009] describe EMA, also based upon the integration of Soar and cognitive appraisal, focusing on the separation of (slower) cognitive process to represent the situation and (faster) appraisal processes. The coping mechanisms include uncritical reasoning, such as "mental disengagement" (as in coming to care less about a goal that has a low probability of success) as well as critical reasoning, such as "seeking information". However, the appraisal mechanisms are not derived upon theories about human behaviour in a group context; for example there is no coping mechanism to avoid damage to self esteem by ignoring information. [Laird 2008] describes an "Appraisal Detector" in Soar that implements cognitive appraisal theory as a sub-symbolic extension to the core Soar system, potentially allowing all cognitive functions, such as memory, to be affected by emotion, but in practice only supporting the learning function.

Other cognitive architectures have been used to model emotions. [Fum & Stocco 2004] uses ideas from ACT-R [Anderson et al 2004] to model the effects of emotion on human performance in a specific task, based upon the changing of the ACT-R memory activation mechanism to include emotional aspects. However the calculation of the emotional aspects is specific to the domain of the task and is not a generic mechanism based on a theory such as cognitive appraisal; nor does the research concern itself to the differences between critical and uncritical reasoning. [Lin et al 2011] describes EmoCog, a proposed cognitive architecture that includes an emotion component based upon appraisal theory, where memory nodes have associated emotional values that can focus attention (and hence cognitive processing) onto the most salient events and information. This work does discuss the effects of emotion on the fidelity of the reasoning when executing a plan, where logically necessary steps are omitted due to low emotional arousal, but no specific appraisal mechanisms are given.

A different approach to the modelling of cognitive agents is that of BDI (Beliefs, Desires & Intentions), e.g. [Rao & Georgeff, 1995], which focuses on committing to goals in response to beliefs about the world and internal desires, and then detailed planning, and executing actions, to achieve these goals (the combination of commitment and plan being an "intention"). At this level of description, our work is similar to BDI in that we model agent behaviour that maintains beliefs about the world, has underlying desires and performs actions that are derived from satisfying these desires. However the BDI approach is generic and does not, in itself, provide a mechanism for determining desires (and the acceptance of beliefs) from social science theories of appraising and coping in a group context. Furthermore it seems to be taken for granted that BDI is concerned with modelling cognition that aims to correctly achieve specific tasks, rather than simulating the effects of emotion on uncritical reasoning.

The work surveyed above addresses many of our key issues, but none completely address our goal of simulating human behaviour which may consist of reactive responses to situations and may be based upon uncritical reasoning, false assumptions and cognitive biases. Our goal is to faithfully replicate and explain such behaviour, be it good or bad, in terms of the effects of emotion, rather than seeking to create effective problem-solving behaviour for specific tasks.

One body of research that addresses all of our target issues is that of Silverman, for example [Nye & Silverman 2013] describes the application of their cognitive architecture, PMFServ, to social learning. This contains an attentional mechanism based on social and other cues such as authority and influence of in-groups, and a motivation system for actions that is based on cognitive appraisal theory. Potential actions are appraised (via "activations") against a tree of goals, standards and preferences, some of which are social, such as "esteem", "treatment of out-groups" and "desirable future for the group" respectively. The activations are analysed to form a set of emotions which are then combined to form a subjective utility for each action, taking account of the expected change in emotions caused by an action. The action with the highest utility is then chosen for execution (though there are some other constraints that may rule out the choice of an action). Whereas this work addresses our target issues, at a more detailed level there are differences. Social group aspects of self-esteem and in- and out-group membership make up part of their appraisal mechanism, but these are linked in an indirect way via the mathematical formula for subjective utility, and there seems to be no specific symbolic chain of reasoning based upon a social theory such as Social Identity Theory to act as an explicit causal link between group information through the appraisal process to the emotional coping process. This may make it difficult to represent alternative reasoning mechanisms in the individual, such as the change from uncritical (emotional) to critical thinking. More generally, we are researching into the passing of rationale as part of the communication between individuals (as described in the section on trustworthiness below) and this requires explicit representation of the reasoning between inputs, appraisal, making of assumptions, and the resulting behaviour in terms of the memes to be communicated.

3 SCENARIO

A simple initial vignette (i.e. a detailed part of a larger contextual scenario) has been chosen that provides sufficient detail of an environment, individuals and groups, and has the capability of demonstrating mutable group behaviour together with opportunities for monitoring and intervention. This is a key concern of the coalition commander conducting civil-military operations to engage in, influence, or exploit relations between military forces, indigenous populations, and civilian organizations in support of stability and counterinsurgency within a host nation or region [Headquarters, Department of the Army 2013, Headquarters, Department of the Army 2014]. Coalition forces build trust and influence group behaviours by immersing themselves in the local culture and politics, engaging with local leaders and the general populace to develop partner capacity-building programs focused on host nation governance and economic development [Department of the Army 2014]. Soldiers require cross-cultural language skills and an understanding of socio-cultural relationships in order to provide information messages that are culturally acceptable to the local public [Headquarters, Department of the Army 2013].

The initial vignette is based upon the passing of "pronouncements", or rumours, between individuals belonging to different, competing, groups; this allows the study of how "memes" [Dennett 1995] are taken up, spread or rejected by the communities, and how this affects reasoning both critical and uncritical. To provide some context and linking into the overarching scenario, a "back story" has been invented that provides some motivation for the individuals and the competing groups, based upon an ancient conflict between the "Reds" group and the "Greens" group in respect of land rights.

4 BASIC CONCEPTUAL MODEL

We define a conceptual model of the individual, including their set of beliefs, and an emotional "vector" that defines the individual's overall emotional state and the emotional strength of their beliefs. Groups are modelled with their antagonistic and collaborative relations to other groups, and the relationship of individuals to groups are defined in terms of the in-groups to which they identify and the out-groups with which they "un-identify". Communication between individuals is modelled as "pronouncements", where a speaker tries to pass on some information (a meme) to another individual in a face to face meeting.

Beliefs and pronouncements have linguistic semantic content in terms of a simple subject-action-object triple, and this content may be consistent with, or conflict with, the semantic content of other beliefs; for example "eats chocolate" and "bans chocolate" are inconsistent. Simple semantic reasoning is used to determine such consistency. We also define the emotional "content" of words which may be different for different groups.

The cognitive model takes theories from social psychology as a starting point, as outlined below. However a key part of the research is to develop some of the details in order to define a model that is computational, and our interpretation of these theories, as described in section 6, is relevant towards that aim. In some cases, intuitive hypotheses as to possible social effects have been devised. We do not suggest that the cognitive model presented here is complete and fully validated, but we do aim to demonstrate that the techniques described offer a means of representing and exploring different theories.

4.1 Cognitive Appraisal Theory

To construct a cognitive emotional model of the individual, we appeal to "Cognitive Appraisal Theory" [Smith & Lazarus 1990], which proposes how emotion and cognition are interconnected in two stages: appraisal and coping. In the appraisal stage, the current situation (i.e. the individual's relationship to the environment) is characterised across a number of "Appraisal Variables" including relevance and desirability in respect of goals, possible damage to self-esteem, unexpectedness, causal attribution, controllability, and how well the individual can cope with the situation. The values of the set of appraisal variables can then be mapped into a specification of an emotion. Cognition is involved in appraisal, since the world needs to be compared to the individual's beliefs, and inferences about complex factors such as causality are required. In the coping stage, the individual seeks to reduce the negative "damaging" effects of the appraisal; coping may be problem-focused, such as the use of negotiation, planning and logical problem solving; alternatively coping may be emotion-focused, such as the use of denial, shifting of blame or reducing the importance of the damage. Cognition is also involved in coping, since inference may be needed for solving problems and strategies may be needed for denial, etc. These two types of coping, problem-focused and emotion-focused, may be seen as applying critical thinking and uncritical thinking respectively.

4.2 Social Identity Theory

However, Cognitive Appraisal Theory does not state how appraisal is affected by group membership, and it is necessary to employ further social science theories about the relationships between individuals and groups in order to build a more accurate cognitive model. For example, Social Identity Theory [Tajfel & Turner 1979] defines how individuals create a "social identity" which is part of their concept of "self" that is based upon the groups with they identify and un-identify[1]. In this way, it is possible to model group dynamics, in the way that groups are formed and unformed.

5 COGNITIVE ARCHITECTURE FOR SIMULATING COGNITION

5.1 Overview

Since theory suggests that emotion and cognition affect each other, our computational model of cognition in the group context must take this into account. We use a Cognitive Architecture (CA), since this provides a predefined computational framework based upon theories of human cognition, but this must be extended to cover group-individual interactions and how they are affected by emotion. There are several candidate CAs, including ACT-R [Anderson et al 2004] and Soar [Laird 2008], but after review, Soar was chosen. In addition, there is existing research on the integration of emotion to cognition based upon Soar [Marinier & Laird 2007, Gratch & Marsella 2004, Marsella & Gratch 2009], although this does not address the effects on group behaviour.

Our work combines Cognitive Appraisal Theory and Social Identity Theory (including Self-categorisation Theory) to address emotion and cognition in relation to groups. An initial cognitive model has been constructed in the form of Soar rules, based upon the concepts defined in the conceptual model above together with an "implementation" of Cognitive Appraisal Theory and Social Identity theory, where appraisal is principally based upon self-esteem in relation to in- and out-groups, and coping strategies are based upon the limitation of damage to self-esteem. For example, a pronouncement is damaging to self-esteem when it is about an in-group but is inconsistent with the individual's beliefs. Coping with the damage may be undertaken by uncritical thinking emotional strategies such as rejecting the pronouncement or by critical thinking problem-solving strategies such as removing premises (e.g. being a member of an in-group) that lead to damaging appraisals. Choice of strategies is dependent upon the emotional fear level of the individual, high fear leading to uncritical reasoning and low fear level leading to critical thinking. The individual fear level itself is modelled as to be affected by the interactions between individuals, more interactions with out-group members leads to a higher fear level. It is calculated by the appraisal of the potential for physical (rather than emotional) damage to the self.

The cognitive reasoning performs appraisal of each contact with an external speaker in several ways. Firstly the pronouncement itself is appraised by comparing the semantics of the pronouncement with the individual's beliefs, using a simple semantic comparison, leading to an assessment of potential "damage" to self-esteem <u>if it were to be accepted as true</u>. Secondly the pronouncement is appraised to determine if potential "damage" to the physical self might occur, taking account of factors such as the

[1] More detail is provided in the related theory of Self Categorisation Theory [Turner et al 1987] which describes the cognitive process whereby individuals place themselves into categories (i.e. groups). An individual performs such categorisation by accentuating perceived similarities between members of the same category and perceived differences between members of different categories, using dimensions that the individual considers to be correlated to the categorisation. Furthermore, it explains the difference between an individual's "social identity" and a "personal" identity by the level at which the individual is self-categorising, a social identity being generated by a categorisation at a group level. Currently we do not implement a mechanism for accentuating the difference between groups, so do not apply Self Categorisation theory.

emotional state of the speaker [2]. After acceptance or rejection of the pronouncement, reasoning occurs to decide what pronouncement to pass on in the next contact with another individual; this is determined as the accepted belief with the highest emotional value.

The Soar model is described in more detail below, although the exact representation of the Soar rules and facts is not shown; rather the basic logic of the reasoning is described in informal terms.

5.2 Overview of Components

The Soar cognitive model, when performing uncritical thinking, is diagrammed in Figure 1, where rounded rectangles represent data, square rectangles represent reasoning processes and the arrows show how information flows between them:

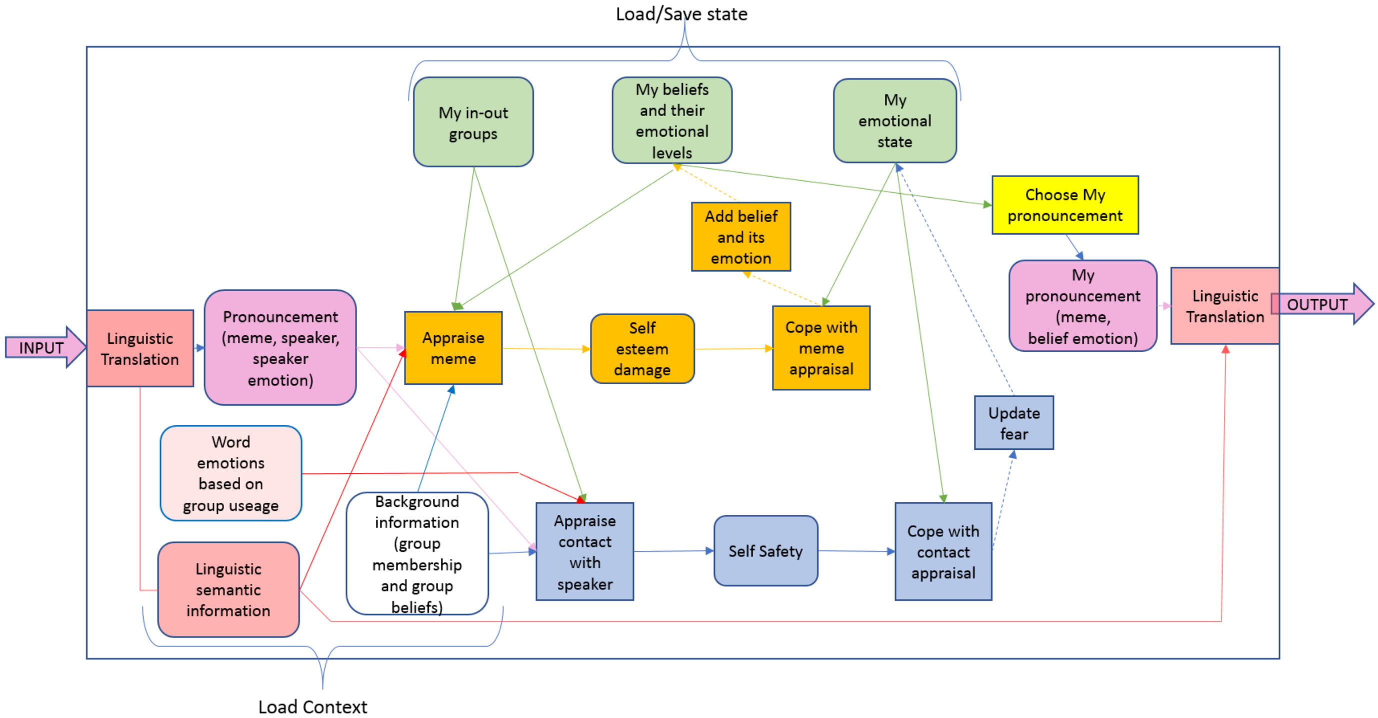

Figure 1 The cognitive model in Soar, for uncritical thinking

The main components of the model are:

- Input/Output of pronouncements
- Linguistic translation and background information
- Linguistic synonyms based upon emotion
- Appraisal and coping of the pronouncement
- Choosing "My" pronouncement
- Appraisal and coping of the contact itself
- Maintenance of "My" state

The additional components involved in critical thinking are highlighted in the centre of Figure 2:

[2] We also model the emotional content of the words in the pronouncement, allowing the possibility that stronger emotive words could lead to a greater potential damage to the physical self.

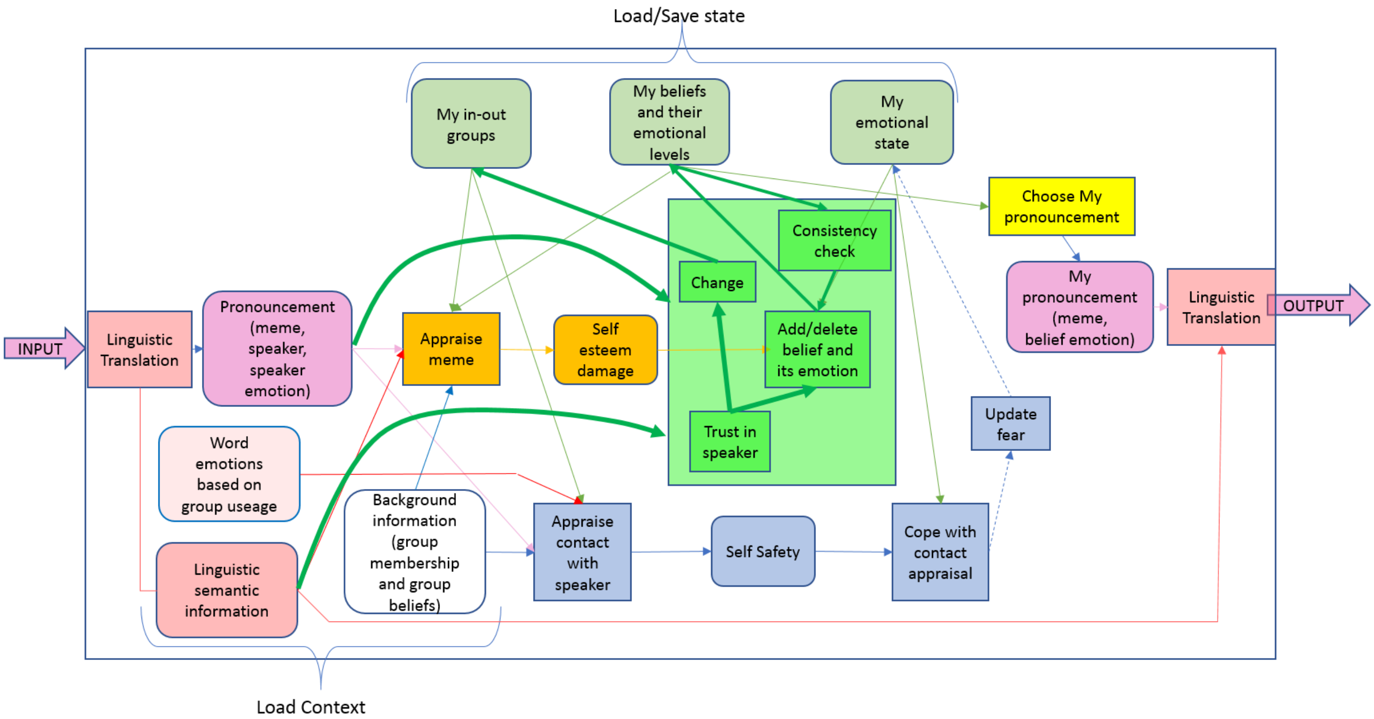

Figure 2 The cognitive model in Soar, for critical thinking

These include:

- The checking, via linguistic semantics, of the consistency of different beliefs
- The deleting of beliefs and their emotions
- The estimation of the trustworthiness of the sources of memes
- The changing of membership of in-groups

Before giving the details of these components, several general concepts will be described.

5.3 Memes

The semantic content of all information to be communicated or believed is represented as a "meme", which contains the following information:

- the "subject", which may be a group, such as "Greens" or an individual, such as "Me"
- the "act", which defines an action, event, or logical relationship involving the subject and object, such as "eats"
- the "object", which may be a group, or individual or other generic concept, such as "chocolate"

Examples of memes (shown as subject, act, object word triplets) are: "Reds ban chocolate", "Me eats chocolate", "Reds are bad".

The actual representation of a meme within the Soar model is not quite as simple as suggested by the above word triplets. This is because it is necessary to define unique identifiers within Soar so that multiple pieces of information are stored against a specific "individual" which can then be used for reasoning. Thus internally, an identifier (such as "M1") is used to represent an individual (such as the person we wish to call "Me") and these identifiers are used within the relationships in Soar's working memory. For this purpose a linguistic translation step is included between the words contained in the input and output pronouncement memes that convert these words to internal identifiers. In the current model this translation is limited, mapping those words that name individuals and groups (such as "Me", "Reds") to identifiers whilst leaving the other words "as is". However it opens up the possibility of more complex (and potentially group-based) linguistic processing.

5.4 Linguistic semantic relations

Two different memes may be analysed to determine if they are consistent or inconsistent with each other by simple analysis of the semantic content of the subject, act and object, and based upon basic semantic relationships between the concepts. Two types of inconsistency are defined. The first is "predicate inconsistency" where the act-object components define inconsistent acts[3], for example

- the act "bans" is defined as being inconsistent with the act "eats", therefore the combination "bans chocolate" is inconsistent with the combination "eats chocolate". Thus the meme "Reds bans chocolate" is predicate inconsistent with the meme "Me eats chocolate"
- the act "hugs" is defined as being consistent with the act "loves", therefore "hugs kittens" is consistent with "loves kittens", and the meme "Greens hugs kittens" is predicate consistent with the meme "Me loves kittens"

The second type of inconsistency is "total inconsistency" where the complete subject-act-object components define inconsistent situations. For example:

- the meme "Reds bans chocolate" is total inconsistent with "Reds eats chocolate" (whereas the meme "Reds bans chocolate" is only predicate inconsistent with "Me eats chocolate", not total inconsistent)

5.5 Beliefs

Groups and individuals have sets of "beliefs", which are defined as:

- a meme, such as "Me eats chocolate"
- the agent (individual or group) that believes the meme (this may not necessarily be the same as the subject of the belief, thus "Me" can believe that "Reds ban chocolate".
- an emotion, being the emotional content of the meme with respect to the agent believer. Thus Me might believe that "Me eats chocolate" with a high degree of happiness (see below)

5.6 Emotion

The cognitive model currently represents six emotions[4] [Ekman, 1992]:

- Happiness (or Joy)
- Fear
- Anger
- Sadness
- Surprise
- Disgust

These are represented in an "emotional vector" which is the ordered set of values for the emotions, each value being in the range 0 (no emotion) to 100 (full emotion). It is debatable whether all possible

---

[3] The term "predicate" here is taken from linguistic practice where the verb-object combination is considered to be a predicate on the subject; this the predicate "bans_chocolate(X)" is inconsistent with the predicate "eats_chocolate(X)".

[4] [Plutchik, 1980] describes a further two, Trust and Anticipation

combinations are valid, but such constraints on the possible sets of values are not enforced. However, at this stage only Happiness and Fear are calculated and are treated as if they were opposites[5], by combining them into a single intermediate value, a "Fear/Happiness level" which ranges from -100 (representing full happiness) to 100 (representing full fear); for example Fear/Happiness of -60 is represented as Fear=0, Happiness=60, and Fear/Happiness of 25 is represented as Fear = 25, Happiness = 0, and the Fear and Happiness levels are never positive at the same time. This intermediate combined value is only used to simplify the implementation of the Soar-based computations described below, and there is no logical requirement to manage multiple emotions as continua; indeed the model actually stores fear and happiness as two separate values.

Emotional vectors may associated with different objects in the model, when that object is considered to have an emotional content. In the current model two types of object have associated emotional vectors:

- a belief, each belief having its own emotional vector
- "Me", representing the overall emotional state of the person being modelled.

An emotional vector has a "total emotional value", which is the summation of all emotional values in the emotional vector. (All values are zero or greater, so there is no offsetting of one emotion against another).

It should be noted that the approach using Cognitive Appraisal Theory does not hold emotions as being the fundamental unit of information; instead the emotions are derived from combinations of appraisal variables, see below.

5.7 Appraisal Variables and Vectors

Cognitive Appraisal Theory states that the external environment is first appraised in terms of a number of criteria that affect the individual and have the potential to raise emotions. Such criteria are called appraisal variables, and are each given a value. The following appraisal variables are currently being determined (together with their range of possible values) [6]:

- self-esteem [-100 for significant benefit, 100 for significant damage]
- physical safety [-100 for significant benefit, 100 for significant damage]

Within the cognitive model, the values of these variables make up an "appraisal vector", and such vectors may be associated with different objects and different aspects of these objects. Thus two appraisal vectors are calculated from each pronouncement, one is an appraisal of the content of the pronouncement (the meme), the other is an appraisal of the contact itself within which the pronouncement was made.

The values of the variables contained in an appraisal vector are converted into an emotional vector, after applying the coping stage, which seeks to resolve any problems raised by the appraisal. The appraisal vector may also contain a "reason", that is a simple symbol that stands in for the cognitive reasoning (implemented as Soar rules) that led to this appraisal. This may be used in critical thinking to further review the logic of the appraisal and coping process, and to pass on reasons as part of the pronouncement, as described below. The vector may also contain additional information that is

[5] This does not match the [Plutchik, 1980] circumplex of emotions where fear is opposed to anger and happiness is opposed to sadness. However in [Marsella & Gratch 2009] positive "desirability" is mapped onto hope and joy, whereas negative "desirability" is mapped onto fear and distress, or anger and guilt if causal attribution is involved. In the group context, we take self-esteem as a "desirable" situation, have followed a similar approach.
[6] Cognitive Appraisal Theory defines more variables, and our future work may add to this list; one interesting (but complex) addition would be causal attribution.

relevant to the type of appraisal. For example, the self-esteem appraisal based upon group membership (described below) will include the relevant group.

5.8 Trustworthiness

When critical thinking, the model aims to assess the trustworthiness of the source (i.e. an agent) of a meme, in order to help decide if it is to be believed. It is taken that memes from a trustworthy source will be believable. An agent receives memes via a pronouncement, and this provides two pieces of information that can assist the determination of trustworthiness of the source, the speaker and the reasons given for the meme. Since memes are passed between agents, we model two ways of how a meme is sourced: via an original trusted source; via another speaker who cites the trusted source as the reason for the meme.

To determine that a speaker is a directly trusted source, the current model considers that anyone who is "neutral" (i.e. does not belong to either the Greens or the Reds) is trusted[7]. To determine that there is an indirect trusted source for a meme, the reasons that are passed with memes are used. Thus a reason for a belief may include the fact that it was stated by a trusted source, hence the speaker is being a proxy for the original source.

Thus there are several stages in the propagation of trusted memes. In the first stage a directly trusted source (e.g. a neutral agent N) pronounces a meme M to another person P1, if the meme M is accepted by P1 the corresponding belief BM will have the trusted source N as a reason. In the second stage, P1 pronounces the meme M to another person P2 with the reason that P1 is proxying for the trusted source N. If P2 accepts the meme M then the corresponding belief BM will have the proxied source N as reason. Thus the meme is spread though those agents who are in critical thinking mode.

5.9 Logic of the Cognitive Model

We describe the cognitive model in more detail, based in part on the social science theories that underpin it (emphasised in bold), and in part on intuitive hypotheses as to the nature of cognition.

5.9.1 Inputs and outputs of pronouncements

The inputs and outputs to the model are pronouncements. On contact with another individual, the other individual's pronouncement is received, including the meme containing the semantic content, the identity of the speaker and the emotion expressed by the speaker. As a result of reasoning, "My" pronouncement (the one that "I" most wish to express) is output, including the meme and the emotion of the belief[8]. As described above, there is a linguistic translation between the pronouncement meme and an internal Soar representation.

5.9.2 Linguistic Semantics and Emotional Content of Words

Linguistic semantics (i.e. the semantic relationships between words as described above) and background information is used to compare the meme in a pronouncement to memes contained in

[7] It is desirable to extend this to the modelling of authoritative, scientific reputation, although there are issues of how any authority is to be accepted by others. In addition, being a member of the same group does not of itself guarantee trust.
[8] This raises the question as to the difference between the emotion of an individual (as per the input pronouncement) and the emotion of a belief (as per the output pronouncement). We believe that the individual has an overall emotion, which is separate from (but perhaps related to) the emotions of the beliefs; this would allow the representation of "undirected anxiety".

"my" beliefs, to determine which beliefs are consistent with the pronouncement and which are inconsistent (predicate or total) with the pronouncement[9].

In addition, "emotional synonyms" are defined, that relate synonyms that have an associated emotional vector to basic words that do not. For example, the basic word "dislikes" has no emotional content, whereas the synonym "bans" has an emotional vector (containing fear). Emotional synonyms can potentially be used to estimate the emotion associated with the meme in a pronouncement, specifically by determining the emotional content of the "act" of the meme. Thus a meme such as "Reds bans chocolate" has a high emotional content, whereas "Reds dislikes chocolate" has a low emotional content.

5.9.3 Applying Cognitive Appraisal Theory to the pronouncement meme

The semantic content (meme) of the pronouncement, after determining its consistency or inconsistency with "my" beliefs, is examined **by applying Cognitive Appraisal Theory in its two stages of appraisal and coping.** Appraisal occurs to create the "self-esteem" appraisal variable, by assessing whether the action of a group (in or out) is consistent or inconsistent with "my" beliefs, **using Social Identity Theory,** that specifies that the concept of self is in part determined by the norms of the in-groups with which "I" identify and the norms of the out-groups with which "I" un-identify. This is interpreted in the following rules:

1. If "my" in-group performs an act that is (predicate) inconsistent with my beliefs then "my" self-esteem is damaged
2. If "my" in-group performs an act that is (predicate) consistent with my beliefs then "my" self-esteem is benefited
3. If "my" out-group performs an act that is (predicate) inconsistent with my beliefs then "my" self-esteem is benefited
4. If "my" out-group performs an act that is (predicate) consistent with my beliefs then "my" self-esteem is damaged

This assessment is determined by checking the subject of the pronouncement (to find the group performing the action) and by checking whether the act and object is predicate inconsistent with "my" beliefs, and results in a value for the "self-esteem" appraisal variable. This variable is associated with the meme in the pronouncement, rather than the overall emotional state of "Me". In effect the self-esteem damage is that which would occur <u>if the pronouncement were to be accepted</u>.

The logical reasoning for an appraisal using rule 1, including Social Identity Theory, is shown in Figure 3:

[9] Currently we do not compare beliefs against each other, presuming that any set of prior beliefs given to the model are already consistent with each other.

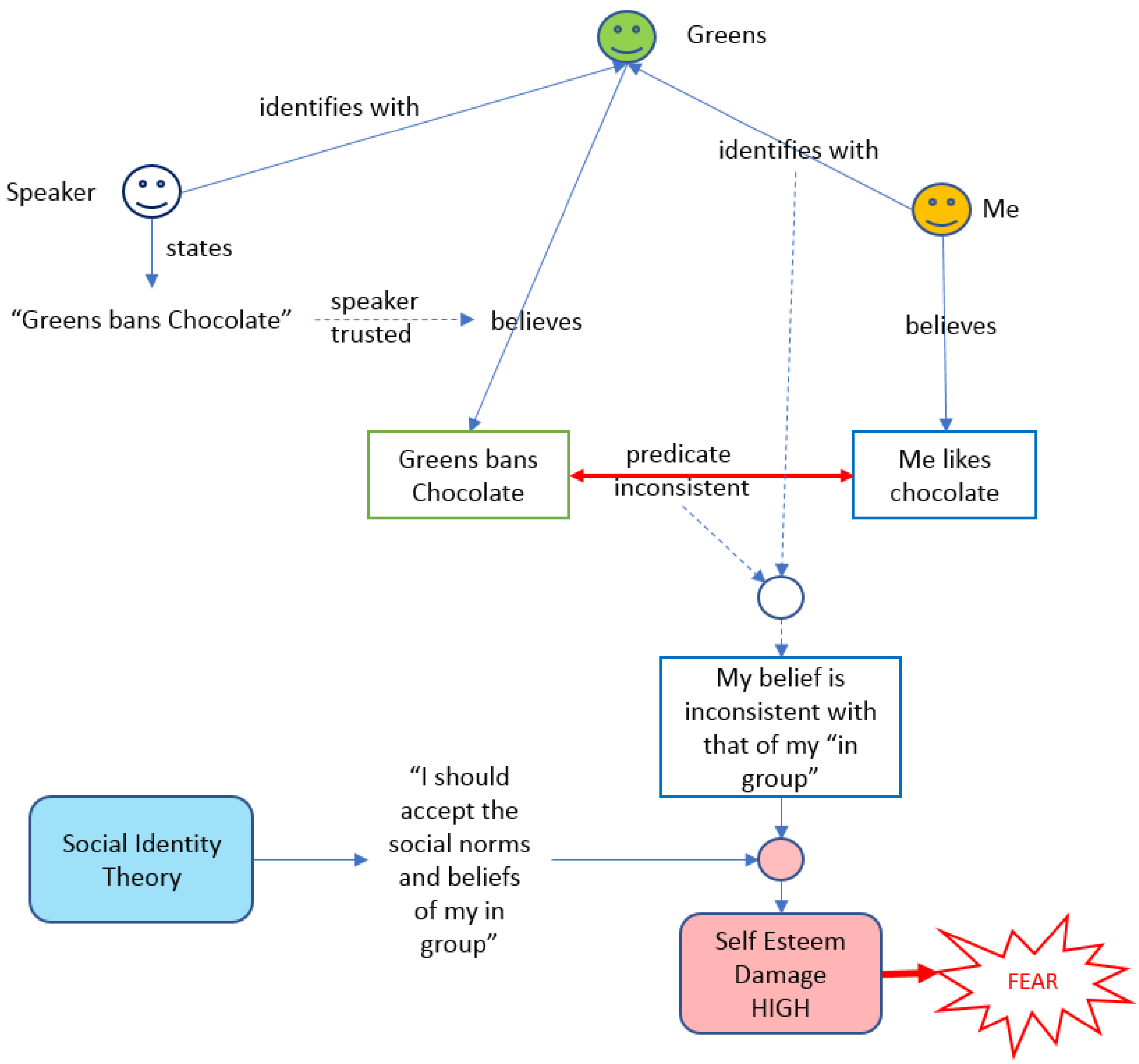

Figure 3 Rationale for an appraisal of high self esteem damage

Here the speaker, who is a "Greens" states that the Greens ban chocolate. Since the speaker is trusteed (because they are a Greens, see below for further discussion on this), it is taken that the Greens do indeed ban chocolate. In contrast "Me", who is also a Greens believes chocolate is likeable, which is predicate inconsistent with the Greens belief. Thus "my" belief is inconsistent with that of "my" in-group, Greens. Social Identity Theory states that one should follow the social norms, including the beliefs, of one's in-groups, but this is not happening, so self-esteem is damaged.

5.9.3.1 Coping and Critical Reasoning

Coping is then performed on the self-esteem damage appraisal. The cognitive appraisal theory suggests that there are two mechanisms for coping, emotion-based and problem solving-based; in the terms of this paper this corresponds to uncritical and critical reasoning respectively. In the model, two factors determine whether the agent is to perform critical reasoning or uncritical reasoning. The first factor is the current emotion vector of "Me", so that critical reasoning only occurs when the agent is at a lower level of fear. The second factor is the influence of the speaker's use of critical reasoning on the listener, which we might call "Critical Reasoning Encouragement"[10], where we hypothesise that an agent might be encouraged to perform critical reasoning (even tolerating a higher level of fear) by the knowledge that the speaker is itself performing critical reasoning.

Both of these factors are implemented in the following rules to determine if the listening agent is to perform critical or uncritical reasoning (and hence problem-solving or emotional coping)

- if the fear level is very high[11] then uncritical reasoning is performed
- if the fear level is high and the speaker is not performing critical reasoning then uncritical reasoning is performed by the listener
- if the fear level is high and the speaker is performing critical reasoning then critical reasoning is performed by the listener
- if the fear level is not high then critical reasoning is performed

Whether the speaker is, or is not doing, critical reasoning may be determined by the cognitive model of the listener by examining whether there are any reasons passed across with the pronouncement. Currently it is assumed that if reasons have been given for the pronouncement then the pronouncement has been arrived at by critical thinking, although this is a logical simplification.

Both types of coping must decide whether to accept or reject the pronouncement and (if accepted) to determine the resulting emotional vector for the belief. (A minor complication is that if the belief is already held by "Me" then the belief is not re-added nor is the existing belief's emotional vector updated). If the belief is accepted then the pronouncement is associated as the "source" of the belief.

In emotional (uncritical) coping, the pronouncement is rejected outright if the self-esteem damage is greater than 0; in this way the damage to the self-esteem is not taken (which would have been if the pronouncement were accepted). Pronouncements are accepted which are appraised with negative self-esteem damage, that way self-esteem is enhanced. In addition, the resulting emotional vector associated with the belief is determined from the self-esteem appraisal variable as follows:

- if self-esteem > 0 then this represents damage and the belief Fear/Happiness = damage
- if self-esteem <= 0 then this represents benefit and the belief Fear/Happiness is = -(damage)

Thus emotion of a specific belief arises only indirectly, calculated from the primary information in appraisal variables.

In some situations, coping also leads to a change in the overall person's emotional state as well as to a change in the emotion of specific beliefs. For example, as described in section 5.9.5, rejection of a damaging belief has a damaging "personal cost" in the calculation of the person's emotional state.

In problem solving (critical) coping, inspection of the reasoning that led to the appraisal can suggest how self-esteem damage may be avoided. In the example above, the damage is evaluated from two premises: that the speaker is to be believed; that "Me" identifies with the in-group Greens. Thus, logically there are two possible ways to remove the damage, by disbelieving the speaker, or by leaving the in-group. Consider the action of leaving the in-group, shown as a crossing out of the relevant premise ("identifies with") in Figure 4:

[10] This is similar (but opposite) to emotional contagion, described below, where the emotion of the speaker affects the emotion of the listener.
[11] The levels that define high and very high fear may be set by the user in the Repast simulation environment described below.

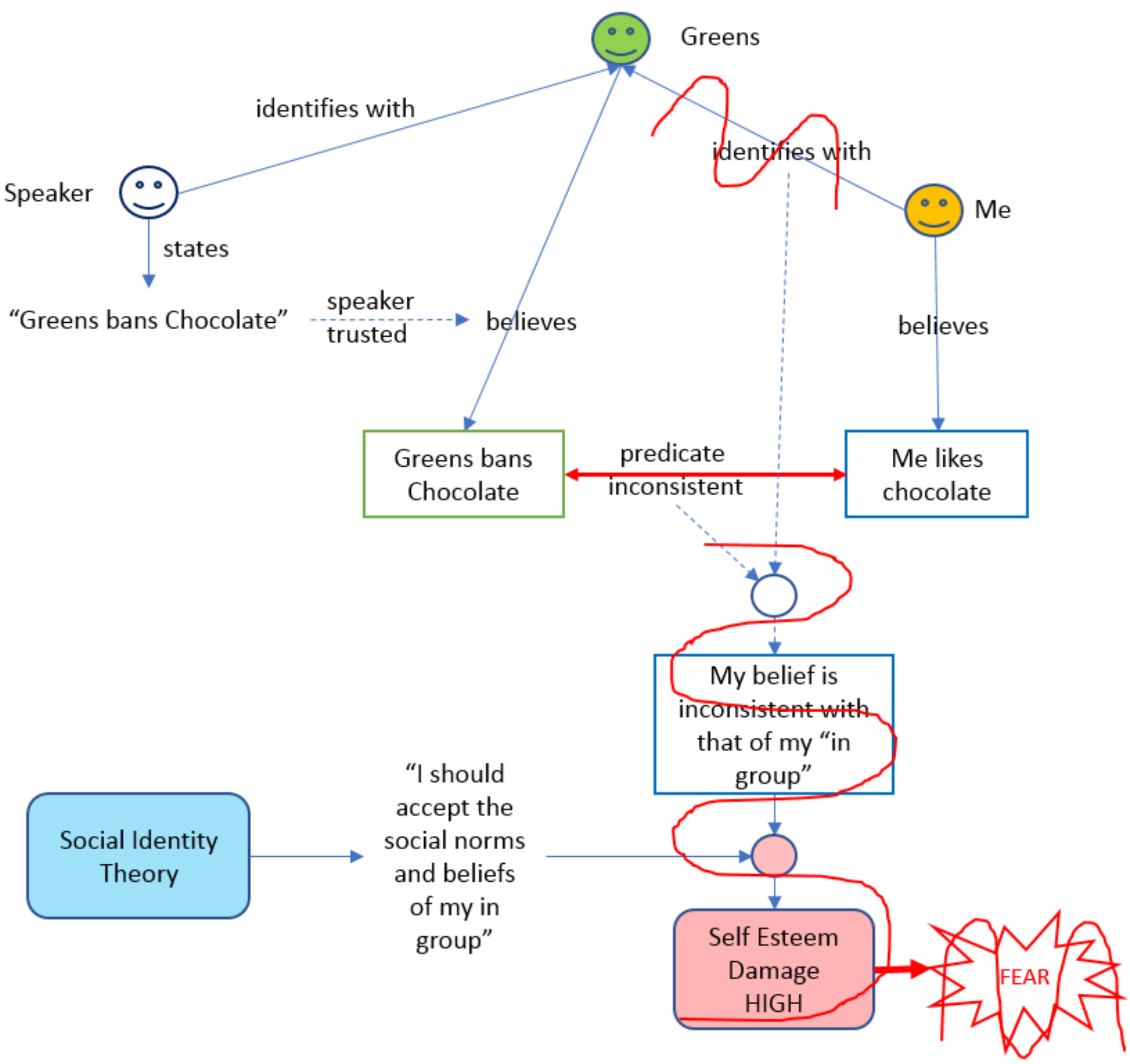

Figure 4 Rationale after removing Greens as an "in-group", thus removing self-esteem damage

This results in the removal of dependent inferences, including the appraisal of high damage (shown as lines crossing out the inferences). This critical thinking coping strategy may be expressed as:

- if the reason for the self-esteem appraisal is that the associated in-group is being stated as having an inconsistent belief and the speaker is trusted, then the group is removed as one of "my" in-groups, and the pronouncement is accepted

This contrasts with uncritical thinking coping, where the pronouncement is just rejected out of hand with no further analysis. Consideration of the diagram shows that this would lead to an inconsistent state of knowledge where a speaker is both believed (because the speaker is trusted) and not believed (because the pronouncement is rejected), but this presumably would not be noticed by the individual, otherwise they might engage in critical thinking to resolve the inconsistency.

Another logical possibility is to not believe the speaker, which would remove the inference that "Greens ban chocolate" leading to the removal of the appraisal from the diagram. This is currently implemented in a different way, by determining the trustworthiness of the speaker prior to making the appraisal, and by only performing the appraisal if the speaker is trustworthy. Therefore no specific coping strategy is needed, since the appraisal has already taken this into account. This shows that there are alternative mechanisms for implementing the logic of appraisal and coping.

At this stage of the research, these logical considerations in support of the appraisal and critical coping strategies have been done "by hand" leading to the design of Soar rules in the model. In previous work, [Mott et al 2010, 2015] we analysed the rationale and automatically calculated the premises and assumptions that led to certain conclusions. But this requires the underlying reasoning engine to have capabilities for examining the rationale graph, and this would require modifications to the Soar inference system. Such modifications could lead the cognitive architecture towards the concepts of argumentation theory [Dung 1995]. Some initial work has been done in the recording of reasons for beliefs and the passing of reasons with the pronouncement itself, in order to model the use of reasons as attempting to convince others of the validity of statements, [Sperber and Mercier, 2017] thus being part of the cognitive and emotional processes.

5.9.4 Determining "My" pronouncement

After the pronouncement has been accepted or rejected, the current set of beliefs is examined to see which has the most emotional salience and therefore is the one that "I" choose to pass on to the contact in return. This choice[12] is made as follows:

- prefer a belief whose total emotional level is higher
- of beliefs whose total emotional level are the same, prefer a belief whose source is a pronouncement[13]

5.9.5 Applying Cognitive Appraisal Theory to determine the person's emotional state

Intuitively, the contents of the pronouncement and the nature of the contact could have an effect on the overall emotional state of the person, as well as on the damage to self-esteem described above. An additional appraisal variable is used for this purpose, that of the potential damage to the person's physical safety (called here "self-safety") as a result of the contact. Such an appraisal must be coped with and may lead to an effect on the person's overall emotional state. There are a number of factors involved in the contact and the pronouncement that could suggest potential damage to self safety as assessed by the individual, and this assessment could be affected by the emotional state of the individual. One factor relates to the emotional state of the speaker, on the grounds of "emotional contagion" where a listener may be "contaminated" by the speaker's emotion; this information is directly available from the pronouncement[14]. Another factor is the potential effect of the semantic content of the meme on the person's emotion,[15] though this can only be derived from the assessment of the meme against the person's beliefs in some way. Various alternative sources of appraisal of self-safety have been considered, and the model is still being developed in this area. The current appraisal of "self-safety" uses the following factors based upon the emotions that can be derived from the pronouncement:

1. the emotional state of the speaker of the pronouncement; a greater level of fear suggests a greater potential of damage to self-safety, whilst a greater level of happiness suggests a reduced potential of damage to self-safety
2. the emotional value of a belief resulting from accepting a pronouncement; a greater level of fear suggests a greater potential of damage, whilst a greater level of happiness suggests a reduced potential of damage to self-safety

[12] It would be possible to change the model so that more than one belief is passed on, leading to different simulation results.
[13] This approximates to preferring more recent beliefs; it may be better to maintain a "recency" value for beliefs
[14] In theory it could also be assessed from the emotional content of the words in the pronouncement meme
[15] If we do not model this connection, then the content of the memes being passed are totally irrelevant to the person's emotional state; not only is this somewhat implausible, it also disallows the possibility that interventions could ever be based upon the propagation of information.

together with a factor derived from the coping process of accepting or rejecting beliefs that was described above, based upon the idea that there is a "personal cost" to rejecting damaging beliefs[16]:

3. the damage to self-esteem, if the rejected pronouncement meme had been accepted

The appraisal value for "self-safety" is thus a combination of these three factors. At the coping stage, if the person is in a state of uncritical thinking, the person's overall emotion is increased by a proportion of this self-safety appraisal value:

- the person emotion is calculated from the old Fear/Happiness level (prior to the contact) + K * the self-safety appraisal.

where K is somewhat arbitrarily set to 0.2. However, if the person is critical thinking, then the appraisal of potential damage to self-safety is ignored, and the person's emotion does not change.

5.9.6 Maintaining "My" state

Various pieces of information held within the model represent the "My" current state and these may change over time, as more contacts occur. Such changing information comprises:

- the groups with which "I" identify (in-groups) and un-identify (out-groups)
- "my" beliefs (including a meme and the emotional state of that belief)
- "my" overall emotional state
- the reasons for "my" beliefs, including trustworthiness

The current implementation of the Soar model does not maintain this information over time, but instead the state is saved after each contact via the Java-based wrapper, and the Soar system is re-initialised with this saved state prior to the next contact[17].

6 INTEGRATION TO MULTI-AGENT SIMULATION

The running of the Soar cognitive model implements an "interaction" between two individuals when a pronouncement is made, leading to the acceptance or rejection of the pronouncement and the readying of the individual to pass on their own pronouncement at the next contact. We have integrated this individual behaviour into a larger scale simulation with multiple agents, to provide a framework for exploring how larger scale behaviour might "emerge" from the individual interactions and exchanging of pronouncements. For this purpose REPAST Simphony [North et al 2013] is used, which provides a platform for developing multiple agents operating in a spatial environment, for the running of simulations and for gathering and analysis of data. We have integrated the Soar Cognitive Architecture into REPAST Simphony, providing a number of separate agents each behaving according to the cognitive model and passing pronouncements to other agents, allowing the running of simulation experiments.

7 EXPERIMENTATION

A simple experiment using the Soar/Repast simulation can be used to demonstrate how the framework can be used to model emotional and uncritical reasoning, and how external monitoring and intervention strategies could be simulated, although this experiment is not intended to be complete and scientifically valid.

The Repast time series graph in Figure 5 shows two alternative "bans" memes ("Reds bans chocolate", "Greens bans chocolate") flowing through two populations of agents, in two different groups

[16] An alternative approach is for the self-esteem coping to accept the damaging belief with a fear level, but to record it is being false so that it is not employed in any of the logical reasoning
[17] This is an implementation detail; logically it should be possible to maintain the state within Soar.

(approximately) 50% Green and 50% Red. These are initially seeded by two pronouncers, one Red and one Green. As per section 5.9.3, these "bans" memes are happy for the "opposite" group and fearful to the "same" group (e.g. "Reds bans chocolate" is happy for the Greens and fearful to the Reds). All agents start with two "like" memes that will be the default pronouncements if the other memes are not accepted ("Me eats chocolate" and "Me likes kittens") These "like" memes are slightly happy to all agents irrespective of the group to which they belong. The initial average fear level of the entire population is 50 (the maximum possible fear level being 100), and the fear threshold below which critical thinking occurs is set so that 67% of the agents start in critical thinking mode.

The graph shows nine lines each displaying the change over time of a particular variable. Numbered from 1 to 9, these lines represent: 1) the overall percentage of the agents that are pronouncing any meme at all, 2) the percentage of agents that are pronouncing the default "likes" meme (e.g. "Me eats chocolate"), 3) the percentage of agents undertaking critical thinking, 4) the percentage of agents belonging to the Red group, 5) the average fear level of all agents 6) the percentage of agents belonging to the Green group, 7) the percentage of agents that have just switched to a different pronouncement meme, 8) the percentage of agents pronouncing the "bans" meme "Reds bans chocolate", 9) the percentage of agents pronouncing the "bans" meme "Greens bans chocolate".

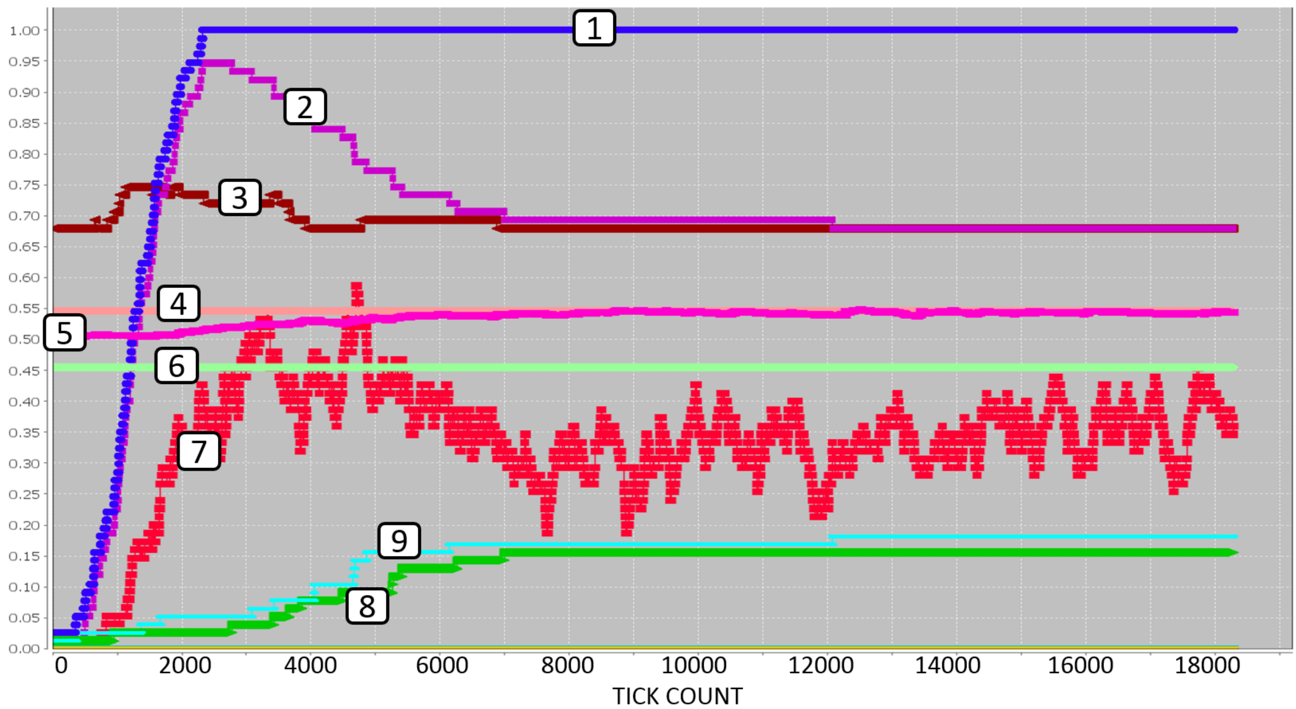

Figure 5 Time series of key variables in a basic Repast Simphony simulation experiment

Several phenomena are visible in this graph and we can relate these to aspects of the model described above. Analysis of the simulation data shows that there are fairly complex interactions between the events, so we only attempt to show some of the main features here.

Up to about tick count 2000, it is mostly the default "likes" memes (2) that are taken up, since there are only two initial pronouncers of "bans" memes, and the critical thinkers will reject the "bans" memes (they are not spoken by a trustworthy source), leading to most agents pronouncing their default meme. Critical thinking encouragement increases the percentage of critical thinkers (3), which ought to inhibit fear increasing. However the small uptake of the "bans" memes that does occur will tend to increase fear (for reasons described in the next paragraph) and this offsets the increase of critical thinkers, causing a small overall increase in fear (5).

From this time on until about tick count 12000, the reach of the "bans" memes increases significantly (8, 9), and thereby eats into the take up of the default "likes" memes (2). As more interactions between the agents occur, "bans" memes are accepted by more non-critical thinking agents (of the "opposite" group) and then pronounced to other agents; some will accept (further increasing the reach) but others (non-critical thinkers of the "same" group as the "bans") will reject the memes. This rejection will increase fear level due to the personal cost for rejecting damaging beliefs (section 5.9.5). Thus the spreading of the "bans" memes will actually cause an increase in fear, and this is seen in line (5). As fear goes up, critical thinking is reduced (3), and the process of critical thinking encouragement is stopped, thus releasing the possibility of "bans" meme acceptance and fear increase.

There is an interesting "micro event" at tick count 12000, where the final acceptance of a "bans" meme takes place (9). This leads to a blip of increased fear (5), and detailed analysis of the data (not shown) indicates that this event causes ripples of increased acceptance and rejection of a bans meme, with its associated increase in fear, over a number of ticks, eventually leading to quiescence. After this, from about tick count 14000 to the end of the run, the values of the main variables do not vary a great deal. No further changes to the percentage reach of the "bans" (8,9) and "likes" (2) memes takes place, and analysis shows that all critical thinking agents hold the "likes" meme and all non-critical thinking (i.e. more fearful) agents hold one or other of the "bans" memes.

These initial simulations are intended to test the framework rather than representing validated scientific results in specific psychological situations. Nevertheless some patterns are beginning to emerge, such as the cyclic swapping of memes, the opposition of positive and negative forces on the fear level and critical thinking levels noted above, and the removal of inconsistent beliefs in the population seeded by a trusted source (not shown). The framework offers a number of parameters that may be varied and which can lead to different group behaviours. Some parameters are contained in the cognitive model (such as the approach to handling self-esteem and self safety and the nature of semantic information about the memes), and other parameters are contained in the simulation itself (such as the make up of the various populations and the groups to which they belong and the memes that many be passed around).

We are extending our experiments to model situations where emergence of relevant and real social phenomena may be hypothesised. Comparing the results of these simulations with the results of real-world studies will serve as a method of validating the models, as well as showing the minimal conditions under which the social phenomena can be observed. For example, we are planning to test the emergence of "false consensus" and “pluralistic ignorance” in these settings. Pluralistic ignorance arises when most individuals in a population privately reject a norm, yet keep their rejection private and conform in public because they misconstrue the public conformity of others as an expression of their private belief. [Bicchieri 2005] explains how non-transparent communication is a condition for pluralistic ignorance to arise. This non-transparency may, for example, be due to fear or bias, which both impede the aggregation of correct information. The current cognitive model incorporates fear, communication and bias, and thus can be used as the starting point for minimal experiments to find test conditions and parameters under which pluralistic ignorance is observed. Relevant parameters include the amount of private belief in the "taboo" meme in the population, the overall fear of the population, the size and number of groups (which affects both fear and bias), and the starting level of critical reasoning in the population. The outcome to be observed is the public pronouncements of individuals: we are interested in cases where there is a significant discrepancy between the empirical distribution of these pronouncements and the distribution of private beliefs in the population, which would be evident of pluralistic ignorance. False consensus arises when an individual assumes that their beliefs are shared by others, e.g. by people in general or by a specific in-group [Ross et al 1976]. To model this, we are extending the model to include the making of assumptions by individuals (for example of what a group believes) and the challenging of such assumed beliefs in order to achieve desirable interventions [Berkowitz 2005].

We also aim to explore additional variables expressed in social psychology theories within the Soar models, such as dynamic social impact theory which provides an explanation for socially transmitted beliefs and development of culture through communication that considers, in addition to the strength, immediacy or proximity, and number of sources exerting social influence, the self-organizing properties of groups, such as spatial clustering and correlation of attitudes, consolidation of minorities, and continuing diversity [Latané, 1996].

8 Extending the cognitive architecture with emotional algorithms

Our current approach is to build the uncritical reasoning aspects on top of the cognitive architecture, as in [Marsella & Gratch 2009], rather than building emotional mechanisms within the architecture itself as in [Marinier & Laird 2004]. However the latter approach has the advantage of modelling the effect of emotion over the system as a whole, involving modules such as memory and attention.

Other researchers have leveraged techniques to handle emotions within cognitive architectures, and it is possible to use these as inspiration for similar techniques within a group context rather than within models of individuals. For instance, [Pirolli 2005] used spreading activation algorithms, which specify the spread of activation strength from one memory to the next, as a model of information scent, and developed models which manipulated the cost-benefit, or utility of the agent moving from one piece of information to the next. [Reitter & Lebiere 2011] used ACT-R memory algorithms as a basis of language evolution, allowing the researchers to model the changes within languages and amongst different groups. [Reitter & Lebiere 2012] used memory decay as a model of information decay within groups, in order to simulate group decision making and social cognition. There are similarities between the base level learning algorithms within ACT-R and the happiness algorithms developed by [Rutledge et al 2014], both of which are time based exponential functions subject to some kind of decay. [Long, Kelley & Avery 2015] successfully expanded the Rutledge model to include additional emotions besides happiness - these additions included fear, anger, sadness, disgust, and surprise. These models allowed for stochastic behaviors to be exhibited by a robot while the robot was executing a navigation task. In order to implement such techniques, it may be necessary to extend the cognitive architecture itself.

9 CONCLUSIONS and FUTURE WORK

We have constructed an initial cognitive model for expressing aspects of group behaviour, based upon social science theories and running on a cognitive architecture, and we believe that the representation of the theories of Cognitive Appraisal and Self Identity combined within a computable cognitive model provides a powerful basis for the explanation and simulation of the effects of emotion on critical and uncritical reasoning.

However, the cognitive model covers only a few aspects of group dynamics and requires extension in a number of areas. Firstly, although the integration of Soar and Repast supports the effects of the group on the individual, in that the group information and group beliefs will be present in each individual simulation, it is also desirable to allow the individual to affect the group (and hence other individuals indirectly), in which case a separate model of the group as it changes over time will be necessary. Such a model could be constructed as a separate Cognitive Architecture with its own group level cognitive model, or could be constructed within Repast providing some group level properties that change over time; in either case a mechanism is necessary to pass the changing group-level information down to the individuals. Secondly, the appraisal could be extended to cover more appraisal variables, such as causal attribution and its effect on perceived damage to self-esteem and the blocking of individuals goals. Thirdly the connectivity of the network could be extended with specific patterns of connection, allowing the exploration of social motifs or geographic-based communications. Fourthly, additional properties and relations amongst groups and individuals could be modelled, including roles of an individual in relation to a group, the status of the group as perceived by the individual and the status of the individual within the group. Fifthly, mathematical calculations could be used to maintain the strength of beliefs based upon recency and degree of exposure, as well as to apply the activation spreading algorithms to influence attentional mechanisms.

We have integrated the cognitive model and architecture into a multi-agent simulation system, allowing the construction of experiments, where parameters may be adjusted to cause different group behaviours. It is important that we construct more detailed and scientifically valid experiments using the principles described above, in order to test the validity of the model and its ability to support the prediction of the mutability of the groups and the effects of interventions. In the long term this research aims to support our understanding of how select information, such as information about health services, veterinary aid, and infrastructure development activities may be conveyed during Military Information Support Operations to favourably influence the local population's attitudes, emotions, and reasoning about civil-military operations and US policy [Headquarters, Department of the Army 2013], as well as understanding how cognitive biases may affect reasoning in the coalition as well as the external groups.

10 ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

11 REFERENCES

Anderson, J. R. & Lebiere, C. (2003) The Newell Test for a theory of cognition. *Behavioral and Brain Sciences*, 26**,** 587-640.

Anderson, J. R., Bothell, D., Byrne. M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004) An Integrated Theory of the Mind, *Psychological Review*, Vol 111(4) 1036-1060

Berkowitz, A.D. (2005) An Overview of the Social Norms Approach, in Lederman, L.C. & Stewart, L.P (Eds.) *Changing the Culture of College Drinking: A socially Situated Health Communication Campaign*, Hampton Press Inc.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

DAIS-ITA (International Technology Alliance in Distributed Analytics and Information Sciences) (2016), https://dais-ita.org

Dennett, D.C. (1995) Darwin's dangerous idea: Evolution and the meanings of life. New York: Simon and Schuster.

Dung, P. M. (1995) "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games." *Artificial Intelligence*, 77: 321-357 (1995).

Ekman, P (1992). "An Argument for Basic Emotions". Cognition and Emotion, 1992, 6 (3/4), 169-200

Fisher, A., (2011) Critical Thinking: an Introduction, Cambridge University Press

Fum, D., & Stocco, (2004) A. Memory, Emotion, and Rationality: An ACT-R interpretation for Gambling Task results, *Proceedings of the Sixth International Conference on Cognitive Modeling*, 172-177, Mahwah, NJ: Lawrence Earlbaum

Gratch, J., & Marsella, S. (2004) A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*(4), 269-306.

Headquarters, Department of the Army (2013). Civil-Military Operations (Joint Publication 3-57). www.dtic.mil/doctrine/new_pubs

Headquarters, Department of the Army (2014*). U.S. Army Functional Concept for Engagement* (TRADOC Pam 525-8-5). http://www.tradoc.army.mil/tpubs/

Laird, J. E., (2008) Extending the Soar Cognitive Architecture, *Proceedings of the 2008 conference on Artificial General Intelligence:* Memphis: IOS press

Latané, B. (1996) Dynamic Social Impact: The Creation of Culture by Communication, *Journal of Communication*, 46(4), 13-25.

Lehman, J.F., Laird, J.E., Rosenbloom P.S., (2006) A gentle introduction to Soar, an Architecture for Human Cognition: 2006 update, *Invitation to cognitive science* 4, 212-249

Lin, J., Sparagen, M., Blythe, & Zyda, M. (2011) EmoCog: Computational Integration of Emotion and Cognitive Architecture, *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference*

Long, L. N., Kelley, T. D., & Avery, E. S. (2015). An Emotion and Temperament Model for Cognitive Mobile Robots. In *24th Conference on Behavior Representation in Modeling and Simulation (BRIMS), Washington, DC*.

Marinier R.P. (2008) *A Computational Unification of Cognitive Control, Emotion, and Learning*, Dissertation, University of Michigan, Ann Arbor

Marinier, R.P., & Laird, J.E., (2004) Toward a Comprehensive Computational Model of Emotions and Feelings, *Proceedings of the Sixth International Conference on Cognitive Modeling*, 172-177, Mahwah, NJ: Lawrence Earlbaum

Marinier, R. P., & Laird, J. E. (2007) Computational Modeling of Mood and Feeling from Emotion. *Proceedings of 29th Meeting of the Cognitive Science Society*.461-466. Nashville: Cognitive Science Society.

Marinier, R.P., Laird, J.E. & Lewis, R.L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research* 10 (2009) 48-69

Marsella, S.C., & Gratch, J. (2009) EMA: A process model of appraisal dynamics, *Cognitive Systems Research*, *10*, 70-90.

Marsella, S., Gratch, J. & Petta, P. (2010) Computational Models of Emotion. In Scherer, K. R., Banziger, T., and Roesch, E. (Eds.), *A Blueprint for Affective Computing: A sourcebook and manual*, Oxford University Press.

Mott, D., Giammanco, C., Dorneich, M.C., Patel, J., & Braines, D. (2010). "Hybrid Rationale and Controlled Natural Language for Shared Understanding". Proceedings of the 6th Knowledge Systems for Coalition Operations, Vancouver, Canada.

Mott, D., Shemanski, D.R., Giammanco, C. ,Braines, D., (2015) Collaborative human-machine analysis using a Controlled Natural Language, SPIE Next-Generation Analyst III, April 2015. Also http://nis-ita.org/science-library/paper/doc-2852

North, M.J., Collier, N.T., Ozik, J., Tatara, E.R., Macal, C.M., Bragan. M. & Sydelko, P. (2013) Complex Adaptive Systems Modeling with Repast Simphony. *Complex Adaptive Systems Modeling* 2013, 1:3

Nye, B.D., & Silverman, B.G. (2013) Social learning and Adoption of New Behavior in a Virtual Agent Society, *Presence*, Vol. 22, No 2, Spring 2013, 110-140

Panskepp, J. (1998) Affective Neuroscience: The Foundations of Human and Animal Emotions, New York, Oxford University Press.

Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive science*, *29*(3), 343-373.

Plutchik, Robert (1980). A General Psychoevolutionary Theory of Emotion, In Plutchik, R., Kellerman, H., (Eds.) Emotion, Theory, research, and experience: Vol. 1. *Theories of emotion*, Academic Press

Rao, A. S., & Georgeff., M.P. (1995). BDI-agents: From Theory to Practice". Proceedings of the First International Conference on Multiagent Systems (ICMAS'95).

Reitter, D., & Lebiere, C. (2011). How groups develop a specialized domain vocabulary: A cognitive multi-agent model. Cognitive Systems Research 12, 175–185.

Reitter, D., & Lebiere, C. (2012). Social Cognition: Memory Decay and Adaptive Information Filtering for Robust Information Maintenance. In *AAAI*.

Russell, J. A. (2003) Core affect and the psychological construction of emotion. *Psychological Review,* 110**,** 145-172.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252-12257.

Smith, C.A. & Lazarus, R. (1990) Emotion and Adaptation. In L. A. Pervin (Ed.), *Handbook of personality: Theory and Research* (pp. 609-637). New York: Guilford Press.

Sperber, D. & Mercier, H. (2017) *The Enigma of Reason: A new Theory of Human Understanding,* Allen Lane.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel. *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole. pp. 33–47.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. (1987). Rediscovering the social group: A self-categorization theory. Oxford: Blackwell