Generalising Social Structure using Interval Type-2 Fuzzy Sets

Christopher K. Frantz[1], Bastin Tony Roy Savarimuthu[2], Martin K. Purvis[2], and Mariusz Nowostawski[3]

[1] College of Enterprise and Development, Otago Polytechnic, New Zealand
christopher.frantz@op.ac.nz
[2] Department of Information Science, University of Otago, New Zealand
{tony.savarimuthu, martin.purvis}@otago.ac.nz
[3] Faculty of Computer Science and Media Technology, Norwegian University of Science and Technology, Norway
mariusz.nowostawski@hig.no

Abstract. To understand the operation of the informal social sphere in human or artificial societies, we need be able to identify their existing behavioural conventions (institutions). This includes the contextualisation of seemingly objective facts with subjective assessments, especially when attempting to capture their meaning in the context of the analysed society. An example for this is numeric information that abstractly expresses attributes such as wealth, but only gains meaning in its societal context. In this work we present a conceptual approach that combines clustering techniques and Interval Type-2 Fuzzy Sets to extract structural information from aggregated subjective micro-level observations. A central objective, beyond the aggregation of information, is to facilitate the analysis on multiple levels of social organisation. We introduce the proposed mechanism and discuss its application potential.

1 Introduction

The seamless interaction of software agents within open environments (i.e. in environments in which actors can join and leave over time) requires the ability to observe and identify behavioural patterns in order to adapt and interpret behaviours that are unknown (i.e. have not been formally encoded at design time) or can change over time. In addition to identifying complex social behaviours, such as composite actions and social interaction patterns, both of which are fundamental characteristics of institutions [12], individuals require the ability to infer information about the social structure of the observed social environment, such as relevant demographic information.

In this work we propose a generic process that allows the generalisation of social structure from observational information. We achieve this by aggregating subjectively categorised micro-level observations on arbitrary level of social organisation, and use Interval Type-2 Fuzzy Sets (IT2FS) to identify patterns of category allocations across ordinally-scaled dimensions.

In Section 2 we outline the motivation for the proposed approach, briefly identify related research fields and existing work. This is followed by a brief introduction to

IT2FS in Section 3. In Section 4 we introduce the essential contribution that consists of a staged use of clustering techniques in addition to IT2FS. Finally, in Section 5, we summarise and discuss application areas for our contribution, but also identify further potential for future work.

2 Motivation and Related Work

When humans interact in new environments, they rely on previous experience to guide their actions. However, to capture the social meaning of actions or interaction patterns (and thus inform their action choice appropriately), they also develop an understanding of the social roles, order and dynamics, in short: the social structure of the social environment they are acting in. Learning about social structure implies a generalisation process in order to make the acquired knowledge transferable to unknown situations and environments. Inferring structural information from observation involves several challenges that apply to humans as much as to artificial entities:

- Bounded rationality [15] – Individuals have a limited ability to keep track of the characteristics of all observed individuals, an aspect that challenges the inference of social structure in open systems.
- Incomplete information – Specific social attributes, such as age, may not be accessible for all observed individuals.
- Locality of observations – Individuals do not have a global view, but are constrained to the observation of their specific social environment.

To deal with those challenges, humans rely on abstraction mechanisms that permit the categorisation of observations (e.g. aggregating individual observations into age groups), while operating subconsciously without relying on the individual's explicit mental attention. Following this rationale, individuals continuously invoke some notion of stereotyping or labelling based on 'implicit social cognition' [6] that aims at categorising observation traces by their structural components. The intrinsic operation further includes the consideration of individuals' biases (e.g. attitudes, self-esteem, previous experience) and situational involvement. Those priming influences shape the interpretation and internalisation of observation traces – making it both product of a situational assessment and subjective influence factor for future assessments at the same time. As a response to personal experiences as well as objectives, individuals can invoke subjective social comparison processes [3] that reflect the relative position or role in a social environment (e.g. allocating oneself in a specific age group), as opposed to capturing a comprehensive objectified picture.

While computational capabilities seemingly permit artificial agents to overcome the challenge of bounded rationality, retaining full information permanently is inefficient, both in terms memory consumption and computational efficiency, especially if attributes change, have only temporary relevance, or require frequent computing. Similar to humans, artificial entities face the challenges of operating on incomplete and local information when attempting to infer social structure.

In this work we propose an initial approach that models stereotyping processes in a generic fashion using Fuzzy Sets [16], or, more specifically, Interval Type-2 Fuzzy Sets

(IT2FS) [17], as the underlying technique. Fuzzy sets represent a natural conceptual fit for the problem of *quantifying ordinal categories for given dimensions* (e.g. Age: 'young', 'middle-aged', 'old'), and are able to *capture the blurry boundaries between those categories*. Moreover, beyond serving as a compatible conceptual mapping for specific category definitions, fuzzy sets can be comprised of multiple individual observations, making them a tool for the analysis of observations on arbitrary levels of aggregation, thus facilitating the identification of social structures on various levels of social organisation, such as micro, meso and macro level. As such, fuzzy sets complement imperfect subjective categorisation processes with analytical facilities to allow the objective characterisation of their aggregated outcomes.

The proposed approach sits at the intersection of norm synthesis, a subfield of normative multi-agent systems, and the social-scientific application of fuzzy sets. Work on norm synthesis include centralised, hybrid, and decentralised approaches. Morales et al. [11] propose a centralised norm synthesiser that monitors agents' behaviours in a traffic scenario and infers and imposes rules at runtime. An alternative hybrid approach by Riveret et al. [14] marries bottom-up norm inference and top-down enforcement in which individual agents play stochastic games and individually nominate a preferred normative strategy based on observed strategy outcomes, which are then put forth as motions, voted on and implemented based on a collective social choice mechanism. Frantz et al. [5] use a decentralised approach in which agents infer generalise behavioural patterns and structural information based on observation and present those in human-readable form using a generic norm representation.

We will highlight related work in the area of fuzzy sets after introducing the underlying concept in more detail in the following Section 3.

3   Interval Type-2 Fuzzy Sets

An essential novel aspect in this work is the use of Interval Type-2 Fuzzy Sets (IT2FS) to facilitate the generalisation and synthesis of non-categorical attributes. Zadeh [16] introduced fuzzy sets as a mechanism to represent uncertain information, the complexity of which he deemed to be in inherent conflict with precision. Instead of unambiguously classifying information as members of well-defined (crisp) sets (as exemplified in Figure 1a), fuzzy sets remove the assumption of unambiguous set associations and instead emphasise a continuous *degree of membership* (with boundaries 0 and 1), reflecting the certainty with which a value is a member of the corresponding fuzzy set (as shown in Figure 1b). This flexibility qualifies fuzzy sets for the use in a wide range of application domains involving classification problems that are characterised by the complexity of input data. Examples of those include micro-controllers [9] and image processing [1], but also social-scientific aspects, such as modelling personality traits [13] or establishing a fuzzy measure of social relationships [7].

Referring to the examples shown in Figure 1, crisp sets (Figure 1a) are characterised by their unambiguous association of input values (here: 3 of the dimension x) with a given set (here: C), with a degree of membership ($\mu_C(3)$) of 1. For the (Type-1) Fuzzy Set K (as shown in Figure 1b) the degree of membership ($\mu_K(3)$) is 0.8.
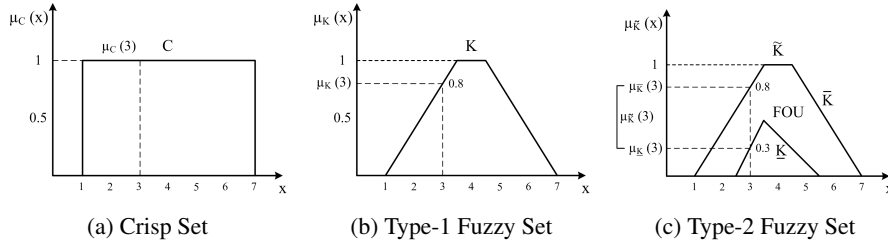
Fig. 1: Examples for Crisp, Type-1, and Type-2 Fuzzy Sets

However, an essential problem associated with Type-1 Fuzzy Sets (T1FS) is the conception of uncertainty as a discrete value, i.e. the representation assumes 'certainty about the uncertainty', here expressed as the degree of membership. A possible solution to this problem is to represent the degree of membership as a fuzzy value itself, making it a recursive problem reflected in Type-n Fuzzy Sets [17] (with n reflecting the order). In this work we concentrate on an interval-based representation of fuzzy sets, specifically Interval Type-2 Fuzzy Sets. In this concept second-order uncertainty is expressed as a *Footprint of Uncertainty* (FOU) that is delimited by an *upper membership function* (UMF) and *lower membership function* (LMF). Consequently, the degree of membership for a specific input value is represented as an interval itself. For the exemplified Type-2 Fuzzy Set $\widetilde{K}$ in Figure 1c the degree of membership for the input value 3 is $\mu_{\widetilde{K}}(3) = [0.3, 0.8]$, determined by the input value's intersections with UMF ($\overline{K}$) and LMF ($\underline{K}$).

More than reflecting a philosophically more accurate representation of uncertainty, the use of IT2FS lends itself well for the representation of systems in which the global state is an emergent property of its constituents' interactions. This enables the inspection on multiple levels of analysis, with the emerging FOU being an essential construct to quantify aspects such as social coherence – an aspect that we exploit in this work. To our knowledge IT2FS have found limited application for the purpose of social modelling, with the exception being their use to quantify the concept of normative alignment [4].

4   Generalising Social Structure

To showcase the use of IT2FS to infer information about social characteristics, we introduce a set of assumptions about the structure of observable information. We assume that each individual carries attributes or markers of numerical (or at least ordinally-scaled) nature, such as their specific age. The corresponding attribute (e.g. 'age') must furthermore apply to all individuals, i.e. each individual must have an age. Individuals must be able to perceive such marker instance values, either based on public display or some form of inference on the part of the observing individual (e.g. gauging another individual's age relative to oneself). Individuals reduce cognitive load (see Section 2) by allocating their observations in ordinally-scaled categories of given dimensions (such as the categories 'low', 'medium', and 'high' for the dimension 'wealth'). These categories are then expressed as value intervals. Following the motivation of this work,

intervals are shaped by the individuals' experiences, with interval centre values bearing higher certainty values than boundary values.[4] However, the proposed generalisation approach is agnostic about the origin of those value intervals.

The devised process exploits the strength of IT2FS to systematically combine individual varying value ranges on arbitrary levels of aggregation. However, to apply IT2FS generation to social systems, we need to take the potential conflicting analytical objectives of both into account. The IT2FS generation process (which we explore in more detail at a later stage) aims to produce a *coherent single membership function* that describes a category of interest while attempting to produce a minimal FOU (i.e. minimising uncertainty about the fuzzy set boundaries) by applying statistical corrections in order to isolate (presumed) irrelevant data or outliers as shown in Figure 2 (individual input intervals are represented in grey colour; the bold red trace reflects the UMF; the green trace represents the LMF). However, social systems exhibit, if not promote, *broad stratification or even polarisation of observed characteristics* (as seen for the clustered intervals on the far left in Figure 2) such as individual or social markers, attitudes, and opinions. Coercing those into a uniform macro-level construct in order to increase coherence (by filtering outliers, etc.) would prevent the comprehensive representation of the existing social landscape and limits explorative analysis, and thus rendering the application of the otherwise appropriate mechanism questionable.
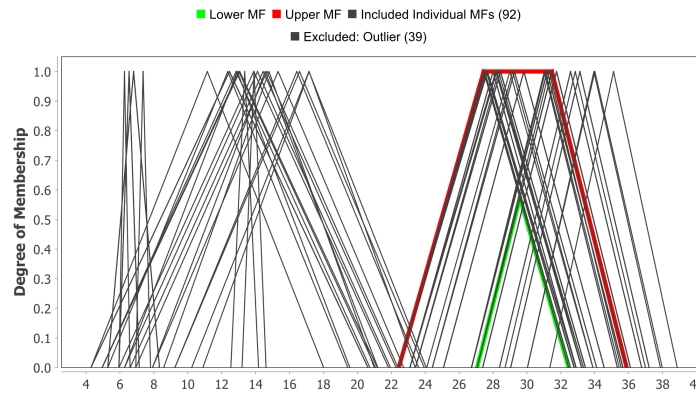

Fig. 2: Exemplified Operation of MF Generation on widely-spread Input Intervals

We thus devise preliminary steps that adapt the use of fuzzy sets for the purpose of social systems by preempting MF generation with steps for both supervised and unsupervised identification of relevant social clusters. The complete process involves

 – the collection of individuals' interpretations of categories for given dimensions, i.e. the numeric intervals describing specific categories (e.g. low, medium, high) within given dimensions (e.g. wealth),
 – the identification of interval clusters for given categories (intra-category clustering),

---

[4] For example, 'low wealth' could comprise the interval 0 to 50 ([0, 50]), medium wealth [50, 100], and anything above 100 considered as 'high wealth'. The membership of boundary values (e.g. 50) with specific sets is of lower certainty than interval centres values (e.g. 75).

– the clustering of interval clusters across all categories within a given dimension (inter-category clustering), and finally,
– the generation of the IT2FS.

We conceive two mechanisms to group this functionality, with the first three steps being managed by the *Clustering Module*, and the remaining ones by the *IT2FS Module*.

Figure 3[5] schematically visualises the overall process. The process is initiated by the injection of collected action observations, and ultimately produces IT2FS membership functions, allowing its interfacing with agent architectures that produce the inputs, and coordination mechanisms that consume the generated membership functions (e.g. to model collective decision-making).
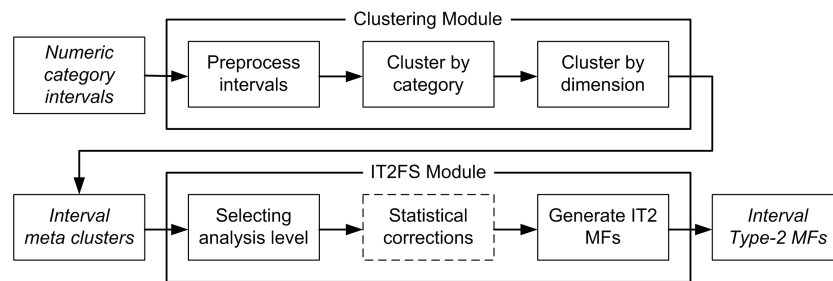

Fig. 3: Process Overview

4.1 Collecting Intervals

As an initial step, category intervals are collected by the interval preprocessor. Assuming the potential operation in open systems, associated tasks involve the sanitisation of input by testing for invalid intervals (such as inverted interval boundaries, infinite or null values). Sanitised intervals are organised by dimension and corresponding categories.

4.2 Clustering Intervals

Clustering (Intra-category Clustering) As exemplified in Figure 2, the identification of a unique set of intervals based on conventional statistical operations cannot accommodate widely-spread input intervals. Instead, we apply density-based clustering in order to identify grouped intervals that may be indicative for a shared conceptual understanding of a given term, i.e. the varying conceptions of 'low wealth' between different individuals (intra-category clustering). To allow the unsupervised identification of clusters we rely on the DBSCAN [2] algorithm that operates on the principle of identifying core points that have at least a specified number of neighbouring points (*minPts*) within a maximum permissible distance ε. For clustering operations explored in the experimental evaluation we consider three as the minimum number of members (*minPts*) to

---

[5] Grouped boxes indicate system components along with elementary processing steps, whereas individual boxes with italicised labels signify processing artefacts such as inputs (e.g. intervals) and outputs (e.g. membership functions). Dashed boxes indicate the optional nature of the operation (e.g. statistical correction).

constitute a cluster. As distance metrics for intervals we use specified minimal range intersections of intervals, with 0 indicating complete overlap of interval ranges (i.e. either identical interval ranges or one range encompassing the other), and 1 indicating no interval overlap. A distance or ε value of 0.3 would thus imply a minimum proportional overlap of 0.7 to consider two intervals clustered.

Meta-Clustering (Inter-category Clustering) The clustering of intervals occurs independently for individual categories in order to characterise varying interpretations for specific categories (e.g. 'low wealth'). However, the individual categories (e.g. 'low', 'medium', 'high') of a given dimension (e.g. 'wealth') do not exist in isolation if we want to characterise social clusters based on their conceptual understanding. We thus perform a meta-clustering operation to integrate the understanding across all category clusters as indicated in Figure 4. Exemplified interval clusters are identified by colour; the horizontal lines highlight the cross-category relationships.
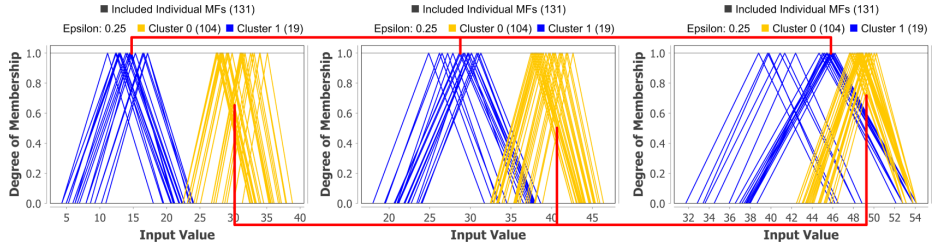

Fig. 4: Meta-Clustering across Categories

The meta-clustering step is based on the assumption that all *individuals hold conceptions across all categories* of a given domain. However, that does *not* imply that individuals with similar conceptions within a given category need to maintain those across all categories of a given dimension. To facilitate the identification of inter-category clusters, each interval *iv* maintains a reference to the originating individual iv_{orig}. In an effort to reduce the number of meta clusters (for larger number of clusters), we devise an optional algorithm. As a first step, all possible cluster combinations across all categories are identified. Following this the proportional intersection of individuals linked to the clustered intervals (relative to the mean size of combined clusters) is determined (with $\{iv, \ldots\}$ as individual category clusters):

$$x_{combination} = \frac{count(\cap(\{iv_{orig}, \ldots\}_1, \ldots, \{iv_{orig}, \ldots\}_k))}{\mu(count(\{iv, \ldots\}_1), \ldots, count(\{iv, \ldots\}_k))} \tag{1}$$

The combination with the *largest proportional intersection* for each cluster is the most representative meta cluster for a given individual cluster.

4.3 Membership Function Generation

The identified meta clusters provide an overview of the presumed social structure based on the differentiated generalised interpretation of conceptual dimensions, but do not

make the individual clusters analytically accessible. Here we invoke Interval Type-2 Fuzzy Sets as introduced in Section 3. The essential purpose of IT2FS is to transform the clustered intervals into a uniform representation that generalises the certainty with which a given input value for a dimension is associated with a category.

Levels of Analysis The process of generalising IT2FS is challenged by the trade-off between representativeness (*quantity* of represented intervals) for the entire category – represented by the UMF – and the IT2FS's *quality*, i.e. its ability to extract a shared understanding of the proximate intervals by introducing some level of certainty – represented by a small FOU (i.e. the difference between UMF and LMF). The quantitative notion of *Representativeness* is thus defined as

$$Representativeness := \frac{count(totalIntervals) - count(excludedIntervals)}{count(totalIntervals)} \quad (2)$$

The qualitative notion of *Alignment* is expressed as the relative difference in area under the LMF relative to the area captured by the UMF (with a value of 1, i.e. identical LMF and UMF, representing highest possible alignment):

$$Alignment := \frac{LMF}{UMF} \quad (3)$$

From a sociological perspective this corresponds to the differentiation into macro- and meso-level analysis (with individual intervals reflecting the micro level). Macro-level analysis thus considers all input intervals for a given category, whereas meso-level analysis concentrates on individual clusters. The selection of meso-level clusters depends on the analytical objectives (e.g. the focus on majority or minority groups), such as the selection of *largest*, *smallest*, most *central* or most *extremal* clusters.

Statistical Corrections In addition to the coarse-grained trade-off based on analytical levels, individual clusters can be refined by statistical corrections inspired by Liu and Mendel [8] to remove noise, emphasise central cluster regions, and enforce at least a minimal aligned understanding. Corresponding non-parametric corrections include

- the filtering of intervals that lie outside a given factor of the interquartile ranges, (to emphasise central intervals), and
- the filtering of non-overlapping intervals (to ensure the establishment of a LMF).

Generating Membership Functions At this stage the intervals have been selected based on analytical strategy and potential further statistical corrections. As indicated at the beginning of this section, we assume that the individual intervals themselves express conceptual understanding of varying certainty, with (full) certainty at interval centres and (in our case linearly) decreasing certainty towards the interval boundaries (e.g. because of overlapping interval regions or dynamically changing boundary values).

Based on the input intervals, the UMF $\overline{\mu_S(x)}$ (for an IT2FS S) is determined as the highest degree of membership for each input x, and LMF $\underline{\mu_S(x)}$ as lowest degree of membership for each input x (see Section 3). The corresponding FOU is then determined as the area between UMF and LMF [10], or union of differences between UMF and LMF membership degrees across all values in X, expressed as:

$$FOU(S) = \bigcup_{x \in X} [\underline{\mu_S(x)}, \overline{\mu_S(x)}] \tag{4}$$

Figure 5 visualises the effects of analysis levels on the generated membership function, with macro-level selection shown in Figure 5a, statistical adjustments to macro-level selection shown in Figure 5b (exclusion of intervals outside 1.5 * interquartile range), and the selection of a specific cluster for MF generation in Figure 5c.

(a) Macro-Level Analysis   (b) Statistical Adjustments   (c) Meso-Level Analysis
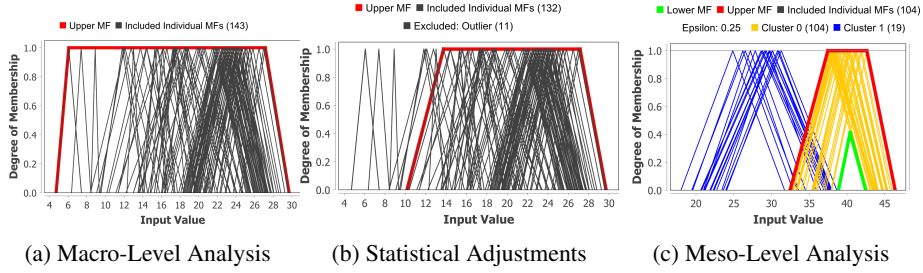
Fig. 5: Configuring IT2FS Generation

The established IT2FSs provide an integrated representation of the chosen intervals with respect to the previously introduced metrics, and furthermore generalise the shared understanding of a given term. This allows the invocation with analytical tools to determine an associated term for a given input across the considered input intervals.

5 Summary, Discussion & Outlook

In this work we have outlined an approach to extract general information about social structures from micro-level observations. This includes the initial identification of category clusters and the subsequent generation of IT2FS membership functions. The presented approach is generic and makes few assumptions about the underlying individuals, which include their *ability to represent observations in a uniform structural representation* and the *ability to subjectively categorise numeric variables*. It lends itself well for autonomous unsupervised operation (only required parameters: granularity of clustering (see Subsection 4.2); choice of desired analysis level and eventual statistical corrections (see Subsection 4.3)). Alternatively, as done in our example, the mechanism can be applied to inspect emerging social clusters and inform a supervised analysis by modifying the configuration (e.g. analysis level) at runtime.

Currently, the proposed approach operates non-intrusively and is only used for analytical purposes. Individual agents neither require awareness nor are they directly affected by their operation. However, looking at future work, the use of fuzzy sets is not constrained to analytical purposes. IT2FS provide a helpful metaphor to instil a computationally accessible mechanism that allows individuals to compare and evaluate their own and others' conceptual understandings. Beyond this, IT2FS can be used to inject notions of computational social choice (such as majority-based decision-making), closing the feedback loop between micro-level entities and emergent meso- or macro-level phenomena.

References

1. T. Acharya and A. K. Ray. Fuzzy Set Theory in Image Processing. In *Image Processing: Principles and Applications*, pages 209–226. John Wiley & Sons, Hoboken (NJ), 2005.

2. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In E. Simoudis, J. Han, and U. Fayyad, editors, *Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, 1996. AAAI Press.

3. L. Festinger. A Theory of Social Comparison Processes. *Human Relations*, 7(2):117–140, 1954.

4. C. Frantz, M. K. Purvis, B. T. R. Savarimuthu, and M. Nowostawski. Analysing the Dynamics of Norm Evolution using Interval Type-2 Fuzzy Sets. In *WI-IAT '14 Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 3, pages 230–237, 2014.

5. C. K. Frantz, M. K. Purvis, B. T. R. Savarimuthu, and M. Nowostawski. Modelling Dynamic Normative Understanding in Agent Societies. *Scalable Computing: Practice and Experience*, 16(4):355–378, 2015.

6. A. G. Greenwald, M. R. Banaji, L. A. Rudman, S. D. Farnham, B. A. Nosek, and D. S. Mellott. A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109(1):3–25, 2002.

7. S. Hassan, M. Salgado, and J. Pavón. Friendship Dynamics: Modelling Social Relationships through a Fuzzy Agent-Based Simulation. *Discrete Dynamics in Nature and Society*, 2011, Article ID 765640:19 pages, 2011.

8. F. Liu and J. M. Mendel. Encoding Words Into Interval Type-2 Fuzzy Sets Using an Interval Approach. *IEEE Transactions on Fuzzy Systems*, 16(6):1503–1521, 2008.

9. Z. Long, Y. Yuanc, and W. Long. Designing fuzzy controllers with variable universes of discourse using input-output data. *Engineering Applications of Artificial Intelligence*, 36:215–221, November 2014.

10. J. Mendel, R. John, and F. Liu. Interval Type-2 Fuzzy Logic Systems Made Simple. *IEEE Transactions on Fuzzy Systems*, 14(6):808–821, Dec 2006.

11. J. Morales, M. López-Sánchez, J. A. Rodriguez-Aguilar, W. Vasconcelos, and M. Wooldridge. Online automated synthesis of compact normative systems. *ACM Trans. Auton. Adapt. Syst.*, 10(1):2:1–2:33, Mar. 2015.

12. D. C. North. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, New York (NY), 1990.

13. T. Ören and N. Ghasem-Aghaee. Personality Representation Processable in Fuzzy Logic for Human Behavior Simulation. In *Proceedings of the 2003 Summer Computer Simulation Conference, Montreal, Canada, July 20-24*, pages 11–18, San Diego (CA), 2003. SCS.

14. R. Riveret, A. Artikis, D. Busquets, and J. Pitt. Self-governance by transfiguration: From learning to prescriptions. In F. Cariani, D. Grossi, J. Meheus, and X. Parent, editors, *Deontic Logic and Normative Systems*, volume 8554 of *Lecture Notes in Computer Science*, pages 177–191. Springer, Springer, 2014.

15. H. A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.

16. L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8(3):338–353, 1965.

17. L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning - I. *Information Sciences*, 8(3):199–249, 1975.