Define Me: A Cognitive and Computational Approach to Critical Digital Identity Representation in Social Networking Applications

D. Fox Harrell
Georgia Institute of Technology
fox.harrell@lcc.gatech.edu

Daniel Upton
Georgia Institute of Technology
dupton3@gatech.edu

Ben Medler
Georgia Institute of Technology
benmedler@gatech.edu

Jichen Zhu
Georgia Institute of Technology
jichen.zhu@lcc.gatech.edu

Invoking a theoretical framework situated at the intersection of humanistic accounts of social identity construction, cognition linguistics research, and digital media technologies, the Advanced Identity Representation (AIR) Project develops theory and technology for users to represent complex, dynamic social identities in digital media such as virtual worlds and social networking sites. Here, we primarily present *DefineMe – Chimera*, a social networking application that uses a dynamic system of categorization and allows users to define each other through metaphorical projection. *DefineMe* is grounded in an interdisciplinary approach that articulates the shared socio-cognitive substrates beneath user representations ranging from user created profiles on social networking sites to avatars in virtual worlds. Secondarily, we present *Identity Share*, a social networking project developed using the *DefineMe* database structure that allows users to define identity categories, share profiles, and anonymously follow each other's web searching paths. The result of the projects is an early articulation of a spectrum of new user identity representations with foci upon group membership, utilization/creation of boundary infrastructures (Bowker & Star 1999; Lave & Wenger 1991), along with cognitive models of metonomy, metaphor, and visual imagery. (Hutchins 1996; Lakoff 1987)

1. Introduction

The Advanced Identity Representation (AIR) Project is the name given to the research endeavor in Fox Harrell's Imagination, Computation, and Expression (ICE) Lab/Studio investigating technology and theory to enable digital experiences that engage a richer range of social identity experiences than those found now in social networking, gaming, and virtual worlds software. We present *DefineMe – Chimera* and *Identity Share* as early steps toward this end. *DefineMe – Chimera* is a social networking application with a novel database system and a front-end Facebook web application. Users can label each other using self-defined predicates expressing their metaphorical similarities to various animals. These descriptions are used as a basis to construct and reconfigure categories on-the-fly as the database grows and to present chimera-like avatar characters to represent the user as composites of various iconic animal graphics. Though this project develops a whimsical, metaphorical model of user representation, the theoretical and technical underpinnings address issues such as coconstruction of identity categories between individuals, marginalization and centrality within identity categories, and the imaginative nature of identity in race, ethnicity, and gender critical contexts.

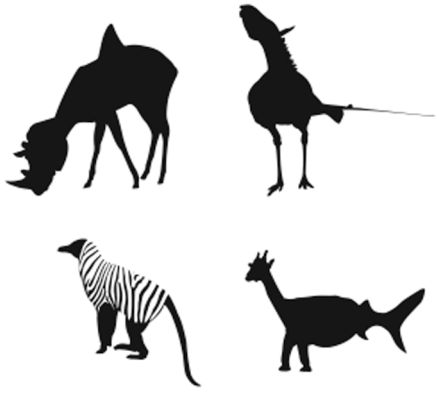

Fig. 1 Metaphorical animal blend avatars potentially generated by DefineMe – Chimera

As a second step toward enabling a new genre of digital media identity experiences, we present *Identity Share*, a critical web-based application that offers a balance between allowing users to author profiles with both self-defined and normative social categories, at the same time as allowing users to specify the relative importance of particular categories. *Identity Share* allows users to anonymously give others permission to follow their web searches, view their wishlists with various websites, and leave comments on their experiences of "(web)walking in another's shoes." The goal is not to connect users as

friends, but rather to allow users to have the uncanny experience of viewing and sharing aspects of each others' needs, values, and desires. Together, *DefineMe: Chimera* and *Identity Share* exemplify early prototypes of the direction that AIR Project systems may take in tackling social identity phenomena in the future.

2. Theoretical Framework

The AIR Project draws on a hybrid approach to issues of categorization and classification. Central influential theories important to the AIR Framework are described below.

2.1 Cognitive Categorization, Metaphor, and Blending

The AIR approach has some of its roots in grounded in cognitive science theory (Lakoff 1987) which asserts that categorization is a matter of both human experience and imagination. George Lakoff's work in this area over two decades ago is well known and influential, yet to our knowledge it is a thread that has been underdeveloped with respect to issues of social identity construction in the critical modes robustly developed in cultural studies with the humanities, especially this approach has not been significantly applied to cases of digital identity representation (an exciting exception being the work of Otto Santa Ana on metaphorical bias in *Brown Tide Rising* (Santa Ana 2002)). Cognitive science research reveals categories as being (1) based on "the same neural and cognitive mechanisms that allow us to perceive and move around" (Lakoff & Johnson 1999), (2) distributed across members of a social group, external artifacts, and even time (Hutchins 1996, 2000), and (3) always situated in particular social and cultural contexts (Lave & Wenger 1991). Important for the purposes here, Lakoff describes a conceptual metaphor-based theory of how imaginative extensions of "prototype effects" result in several phenomena of social identity categorization that are useful for the AIR Project (Lakoff 1987). These phenomena include *representatives* (prototypes) or "best example" members of categories, *stereotypes* that indicate normal, but often misleading, category expectations, and more. Conceptual blending theory builds upon Gilles Fauconnier's mental spaces theory (Fauconnier 1985), elaborates insights from metaphor theory (Fauconnier 2006), and attempts to account for a wider range of semantic phenomena.

2.2 Sociology of Classification Infrastructures

The AIR Project is influenced by accounts of classification from sociology and science studies. In *Sorting Things Out* (Bowker & Star 1999), Geoffrey Bowker and Susan Leigh Star make the case that classification systems are necessary for information exchange and communication. The social challenges regarding classification systems arise from cases where tension exists between contexts, for examples, when one's self-conception differs from prevalent social stereotypes. Important tools for bridging between communities are "boundary objects," defined by Bowker and Star as objects that "inhabit several communities and also satisfy the informational requirements of each." The AIR Project develops what Bowker and Star term "boundary infrastructures." These are defined as "stable regimes managing multiple boundary objects, allowing the necessary information to be accessed by multiple communities." Also crucial from Bowker and Star, is the concept of "torque," the condition where biographies are "twisted in classification systems" to arrive at painful lived experiences. One poignant example Bowker and Star present is the schism between societal and self-perception and the disruptive movement between or misapplication of categories, especially for people labeled as "black" or "colored," in apartheid South Africa. The gap between self (or local community-based) definition of an individual's place in a classification system and hegemonically imposed definition of classifications, and disarming the negative results often arising from such phenomena, is a central to the critique performed by the systems highlighted in this paper. As opposed to computational identity applications that are based on standard, static classification systems, the dynamically configurable, imaginatively grounded AIR Project identity systems are boundary infrastructures that allow users to customize their user profiles and preferences for particular communities.

3. The AIR Framework

Based upon the cognitive and infrastructural approach above, and previous work in imaginative computational discourse and identity construction (Harrell 2007, 2008a), a brief summary of key aspects of the AIR Project's new constructs for implementing and analyzing computational identity follows.

3.1 Shared Technical Underpinnings of Identity Applications

A technical infrastructure-oriented means to compare computational identities is the first pillar the approach developed in this project. Various computational identity applications such as social networking sites, avatar creation systems for virtual worlds, and games are implemented using a limited and often overlapping set of techniques. Fig. 2 describes, at a high level, the components that comprise the majority of widely used computational identity technologies (Harrell 2008b).

The six components in Fig. 2 that commonly form the basis for avatar/character/profile construction can en-

able dynamic and contingent models of social identity in digital environments as described in (Gee 2003). Understanding the reciprocities and overlaps between the technical means by which users stage their identities across digital media forms can enable more powerful customizability and cross-community communication facility in social identity systems.

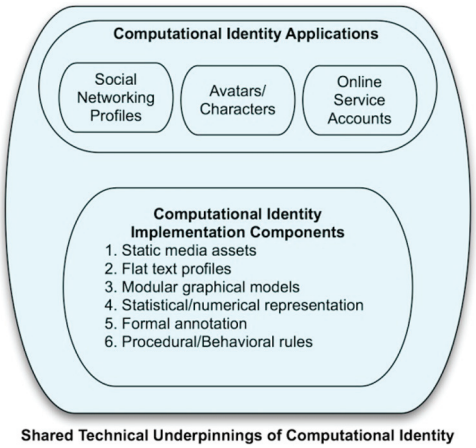

Fig. 2 Shared Technical Underpinnings of Computational Identity Applications

3.2 Cognitively Grounded Model of Computational Identity

The cognitively grounded model of computational identity of the AIR Project is summarized in Fig. 3. It forms the second pillar of our approach.

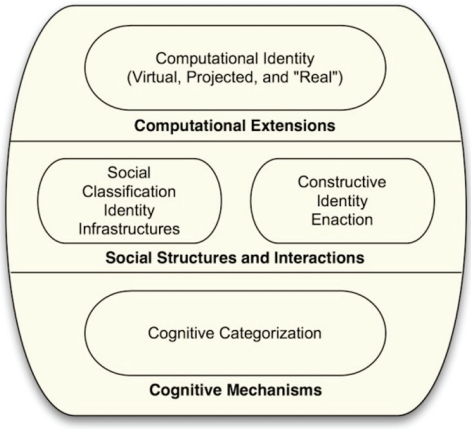

Fig. 3 A Cognitively Grounded Model of Computational Identity

Our digital identity models serve to critique infrastructures of social classification that can, unfortunately, often serve to reify naïve models of identity, and which do no capture the dynamic, constructive, and enacted identity phenomena encountered in everyday experience. This model is an analytical construct used to help to understand the interplay between underlying infrastructures, such as the technical underpinning described above, and the subjective interpretation of digital identities. The utility of the model is that we can identify where schisms exist between a technical structure (e.g., a data structure specifying a player character as a string called a "priest" and an associated "heal" procedure that allows addition to an integer called "hit points) and a real world idealized cognitive model as encoded in a classification infrastructure such as "occupation description" (e.g., the description of a priest perhaps as either someone versed in a metaphysical body of knowledge or as merely the facilitator of a particular type of institution). We can then construct new infrastructures, using techniques such as suggested in the AIR Project, that more closely align these structures and models in order to construct the hybrid of computationally afforded identities and real-world identities that James Gee calls the "projected identity" as shown in cognitively grounded AIR model (e.g., a player taking on the role of a priest in a computer role-playing game and trying to be helpful and supportive to her or his friends). The key here is that our understanding of both computational structures and the ways that users interpret them is based in imaginative cognition processes such as conceptual categorization and blending.

4. *DefineMe – Chimera:* A Critical Identity Construction Social Networking Application

The first system constructed using the AIR theoretical framework is a Facebook application entitled *DefineMe*, the first version of which is called *Chimera*. Specifically, we implement aspects of Lakoff's metonymic idealized cognitive models for categorization to allow users to co-construct their own and others' avatars as boundary objects. (Lakoff 1987) The premise behind *DefineMe* is to allow users to define each others' avatars using both commonplace and abstract metaphors. Users can append metadata to other peoples' profiles to drive dynamic generation of avatar images. The initial content domain consists of animal metaphors that can be mixed-and-matched algorithmically. Animal metaphors are potent entrenched metaphors for human personality. (Turner 1996) (e.g., sneaky weasels or docile sheep), however this animal metaphor-based version is only an initial experiment. The model extends to more directly socially engaged categories such as social scenes, fashions, or movements.

The *DefineMe* database is designed to be lightweight, dynamic, and extensible, while implementing categorical relationships between members. When comparing profiles, *DefineMe* is designed to match lexical items and

logical relations directly, or it can compare the structures of profiles following insights from the analogical structure-mapping engine (SME) developed by Ken Forbus et. al. (Forbus 2001; Gentner 1983) The *DefineMe – Chimera* application reported on here focuses on creating metaphorical projections as described above. The *DefineMe* database relies on tags to create additional descriptors for each category or member. For instance, one user could describe her friend as a 'lion' (which would be the member) because she is 'strong' (which is the tag). Another user could add an additional tag, stating that she is a 'lion' because she is 'carnivorous.' These tags can comprise vertical parent-child links (e.g. a 'lion' is-an 'animal') or horizontal implicit links (e.g. in another user's profile a 'lion' is-an 'Ethiopian symbol,' yet the system may still create a category linked by the concept 'lion'). The fact that users define other users has the potential to both entertain and agitate, regardless it allows for critical inquiry into the phenomena echoing real-world labeling.

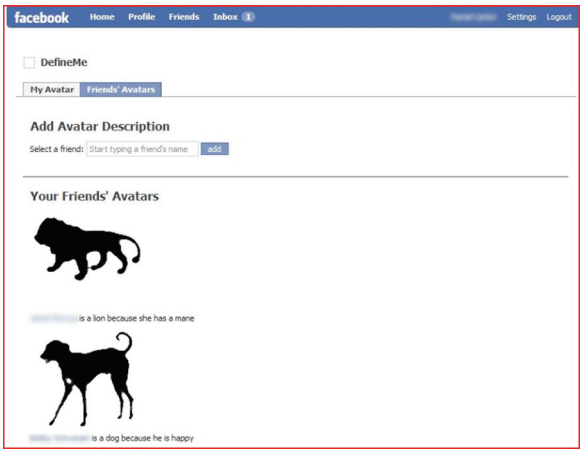

Fig. 4 A screenshot of the DefineMe – Chimera facebook interface

Following the work of Eleanor Rosch cited in (Lakoff 1987), the tagging system can also be used to define aspects of categories themselves. For instance, a 'robin' tag can be added to the category, 'birds,' to define the prototype of that category. In this way, members can belong to multiple groups, but individuals can represent the prototypical members of groups. In this early version, each user is seen as a member of each assigned animal category as well. This membership allows the system to use an individual as a prototype stand-in for the category. For instance, rather than just labeling a friend as a lion, one could state that your friend, Emily, is like your friend Bobby because she is brave. The system can then take all of Bobby's attributes and apply them to Emily's chimera. This relatively lightweight structure avoids some of the pre-defined categorization built into many social networking infrastructures, and has the potential to explore some of the more nuanced identity phenomena mentioned in the theoretical framework above.

5. *Identity Share:* A Critical Identity Construction Social Networking Application

Identity Share, a social networking site for "non-friends," and Daniel Upton's MS thesis project in Digital Media, is also developed under the umbrella of the AIR project. The system allows for social networking by providing users with facilities to construct profiles, follow and comment upon other users, and perform game-like tasks that encourage users to consider exploring both like and different profiles of others. *Identity Share* offers a novel means of self-representation based upon open-ended categories and tags. Standard profile models that typically include normative categories such as name, age, gender, location, and race are replaced with a customizable list that exists as a database, growing as more categories are added. Database consistency is maintained by giving users typeahead functionality when adding custom categories and by presenting existing categories in order from most common to least common. The database structure is based upon the same layout used by *DefineMe – Chimera*. Users can select which categories are most important to them by indicating that they are primary to the user using checkboxes. By allowing for primary selection of categories, we consider the system to implementing centrality phenomena from the cognitive science theory above, i.e., "the idea that some members of a category may be 'better examples' of that category than others," to a users profile. (Lakoff 1987) This means that a user's profile, as a collection of categories that define a user, is no longer viewed as just a set of static characteristics that are true about this user, but rather as a complex set of characteristics where some may be "truer" or more definitional to the user's self-conception. To take this even further, in a future implementation *Identity Share* could offer a ranking system for each category, thereby not only providing centrality, but a centrality gradience, "the idea that members (or sub-categories) which are clearly within the category boundaries may still be more or less central." (Lakoff 1987) This offers a new dynamic to social network profiling that doesn't currently exist on the popular social networking sites.

6. Ethical and Humanistic Implications

When social stakes are low, many people are inclined to reveal their baser selves. Indeed, in a project such as *DefineMe – Chimera* the potential for using the system to ridicule is quite apparent. Likewise, the ability to anonymously follow users' web usage experiences in *Identity Share* offers a potential that may seem to verge on the voyeuristic. Yet, these potentially disempowering uses

are not seen as drawbacks of the systems. Each system is considered to be a culturally situated critical intervention, rather than a usability oriented productive application. In bringing to light more nuanced and imaginative identity phenomena, such as potential ostracism, prejudicial exoticizing of the other, or unflattering labeling, we hope to also provide the potential to disempower such phenomena through dialogic engagement. These systems can be considered cultural productions, or digital media art projects, in the sense that they are provocative cultural interventions situated in an environment increasingly encroached upon by hegemonically enforced, often corporate, models of user identity. As such, the systems succeed only to the degree that they engage users as evocative systems, inspire critical thought, and are construed as adequate for capturing personalities using archetypical avatars or conjure the sensation of experiencing the web through another's eyes. Beyond this, however, we see the systems as prototypes that suggest directions that could enhance the expressive and empowering potentials of productive, utilitarian, or commercial systems such as computer games and popular social networking sites with features such as self-definition of categories and deployment of imaginative metaphor.

Fig. 5 Two screenshots of the Identity Share interface

Despite our provocative and critical interventionist stance, the systems are engineered to mitigate against abuse, and certainly distress of users is not our goal. Looking at the two systems consecutively, mitigating factors designed into the systems are as follows:

DefineMe: Chimera Design Factors

1. Users are only allowed to tag their Facebook "friends" who have added the application.
2. Users have access to a limited database of animal-types.
3. Users must "opt-in" in order to receive a generated avatar.
4. Users can "opt-out" at any time.
5. Users' database entries can be edited by moderators.
6. Users have access to only a limited format for tagging each other.
7. Users can delete entries on their profiles that others have created.

Together, we believe that these factors strongly help to avoid the system's potential to be applied in an overly negative manner. It is a contract between friends to sign up for potential compliments, teasing, and, we believe, self and social insight. Ultimately, *DefineMe – Chimera* is intended to present users with a controlled experience of torqued identity. The fractured identity of a monstrous chimerical representation is then, an accurate reflection of the limitations of applying modular and discrete classifications to a real world biography.

Regarding *Identity Share*, mitigating design factors implemented include the following:

Identity Share Design Factors

1. Users can create their own self-classifications.
2. Users can select which classifications are important to them.
3. Users can avoid or utilize normative categories such as gender or occupation.
4. Users can allow or disallow the system's tracking of their web visitation paths at will.
5. Users' real world identities are kept anonymous.

6. Users' perceived affordances to communicate with one another are highly restricted.

7. Users have full control to delete any of their data in the system.

These factors were developed over the course of iterative refinement of the project based on informal user feedback (mainly via open-ended interviews). The greatest challenge with the system was to allow for user generated categories while also pruning sparsely used and idiosyncratic database elements. A second challenge regarding anonymity and privacy is addressed by careful controls such as articulated above, and by providing quite clear and prominent information on the nature of the site. Quite contrary to being a site to allow people to "spy" on others, it is an "opt-in" site oriented toward users with a desire to share their personal styles, definitions, and web behaviors with one another. Finally, it is a system that is proposed as a balance between the limited and discrete, yet highly modular and structured, information structures provided by digital media and the continuous and transient, yet not computationally amenable, phenomena of identity as shared in the real world.

7. Conclusion

The AIR Project examines the humanistic implications of emerging technologies by seriously considering the cultural effects of user identity within current digital media and the shared sociocognitive foundations that ground their construction. Following various accounts describing the procedural nature of the computational medium (Manovich 2001; Murray 1997), the AIR Project looks at the underlying data structures and algorithms and how they implement cultural identity effects, and posits a technical framework for more deeply engaging identity semantics of classification and categorization. Technologies for implementing socially empowering or expressively critical and transformative experiences are necessary to create experiences to engage real identity social phenomena that lie at the center of so many of our political debates and rich critical fictions.

Acknowledgements

We would like to acknowledge early assistance from the undergraduate researcher Richard Shemaka. Additionally, we thank Carl DiSalvo for significant input on *Identity Share* as Daniel Upton's coadvisor with Fox Harrell.

References

Bowker, G. C. & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*, MIT Press, Cambridge, MA.

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*, MIT Press/Bradford Books, Cambridge.

---- (2006). 'Rethinking Metaphor', in *Cambridge Handbook of Metaphor and Thought*, ed. R. Gibbs, Cambridge University Press, Cambridge, U.K.

Forbus, K. D. (2001). 'Exploring Analogy in the Large', in *The Analogical Mind: Perspectives from Cognitive Science*, MIT Press, Cambridge, MA.

Gee, J. P. (2003). *What Video Games Have To Teach Us About Learning and Literacy*, Palgrave Macmillan, New York City.

Gentner, D. (1983). 'Structure-mapping: A theoretical framework for analogy', *Cognitive Science*, vol. 7, no. 2, pp. 155-70.

Harrell, D. F. (2007). Theory and Technology for Computational Narrative: An Approach to Generative and Interactive Narrative with Bases in Algebraic Semiotics and Cognitive Linguistics, Dissertation thesis, University of California, San Diego.

---- (2008a). 'Algebra of Identity: Skin of Wind, Skin of Streams, Skin of Shadows, Skin of Vapor', in *Critical Digital Studies*, eds A. Kroker & M. Kroker, University of Toronto Press, Toronto.

----(2008b). Digital Metaphors for Phantom Selves: Computation, Mathematics, and Identity in Speculative and Fantastic Fiction and Gaming paper presented to The 29th International Conference on the Fantastic in the Arts, Orlando, FL.

Hutchins, E. (1996). *Cognition in the Wild*, MIT Press, Cambridge, MA.

---- (2000). Distributed Cognition, paper presented to International Encyclopedia of the Social & Behavioral Sciences, University of California, San Diego.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago, IL.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western*

Thought, MIT Press, Cambridge, MA.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate periferal participation*, Cambridge University Press, Cambridge, U.K.

Manovich, L. (2001). *The Language of New Media*, MIT Press, Cambridge, MA.

Murray, J. H. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*, Free Press, New York.

Santa Ana, O. (2002). *Brown Tide Rising: Metaphors of Latinos in Contemporary American Public Discourse*, University of Texas Press, Austin.

Turner, M. (1996). *The Literary Mind: The Origins of Thought and Language*, Oxford UP, New York; Oxford.

Text analysis of large corpora using High Throughput Computing

Mark Hedges
King's College London
mark.c.hedges@googlemail.com

Tobias Blanke
King's College London
tobias.blanke@kcl.ac.uk

Gerhard Brey
King's College London
gerhard.brey@kcl.ac.uk

Richard Palmer
King's College London
richard.d.palmer@kcl.ac.uk

The recent initiative of the US NEH on supercomputing [1] is just one sign that there is a growing interest in the use of highly parallel processing in the humanities. This comes as no surprise if one considers that all over the world, governments and funding bodies are investing heavily in digitization of cultural heritage and humanities research resources. At the same time, the 21st century sciences are demanding an infrastructure to support their advanced computational needs; the computational infrastructure to distribute and process the results from the Large Hadron Collider is just one example. Humanities researchers have therefore begun to investigate these infrastructures to find out whether they can be used to help analyse the extensive, newly available online resources.

We offer an example of such an infrastructure based on High Throughput Computing (HTC). HTC differs from High Performance Computing (HPC), in that the latter relies on hardware specifically designed with performance in mind, whereas the former typically uses multiple instances of more standard computers to accomplish a single computational task. The de facto standard HTC implementation is the Condor Toolkit, developed by the University of Wisconsin-Madison (http://www.cs.wisc.edu/condor). Condor can integrate both dedicated computational clusters and standard desktop machines into one computational resource.

We will present the work of the UK HiTHeR project