Self-Confidence in Autonomous Systems
Papers from the AAAI 2015 Fall Symposium

Believable Character Reasoning and a Measure of Self-Confidence for Autonomous Team Actors

Alexei V. Samsonovich
Krasnow Institute for Advanced Study, George Mason University
asamsono@gmu.edu

Abstract

This work presents a general-purpose character reasoning model intended for usage by autonomous team actors that are acting as believable characters (e.g., human team actors fall into this category). In particular, a measure of self-confidence of an actor in a team is proposed that is based on character believability. The idea is that assuming a cast of believable characters can predetermine a believable (and therefore confident) solution to an unexpected challenge that the team is facing in its rescue mission. This approach in certain cases proves more efficient than an alternative approach based on rational decision making and planning, which ignores the question of character believability. This point is illustrated with a simple numerical example in a virtual world paradigm. Therefore, character believability proves useful and may have a general value for AI.

Introduction

Machine autonomy (Klenk et al., 2013; Roberts et al., 2014, 2015) is vital in heterogeneous (including humans and robots) teams of agents, performing their missions in unexpected challenging situations. This notion of autonomy does not imply isolation from the team, but on the contrary, it means that the actor can act as a useful team member without detailed management or supervision. The sense of self-confidence is one of the keys to this sort of autonomy. In this context, a self-confident autonomous agent is one who can holistically assess robustness of own ability to achieve goals as a part of the team and be useful to the team, **in spite of**, rather than excluding, uncertainties in knowledge and reasoning. A practical measure of this robustness can therefore be used as a measure of self-confidence and as a measure of autonomy, complementing more traditional measures. It is proposed here that this measure can be estimated based on **believability** of character reasoning. Believable character reasoning is a kind of narrative reasoning, when the generated narrative is interpreted in terms of believable characters. Believability of a character means interpretability of its actions in terms of human psychology and human theory-of-mind. The measure simply tells us how consistent are the goals and actions of characters with the attributed to them human-like motives, from the perspective of a given character. In other words, success in actor's interpretation of behavior of self and others in terms of human-like characters implies believability of the actor and suggests its self-confidence. It is argued here that this assumption can greatly facilitate reasoning about other characters in a team involved in distributed team planning and goal reasoning, thereby improving the robustness of the team.

In addition to knowing own domain of affordances (what the agent possibly could do) and own responsibilities (what the agent should do or should not let happen), a self-confident agent needs to determine own domain of competency (what possibly can be done with confidence), and then select goals to pursue and actions to be taken within this domain. The actor's sense of self-confidence based on believability, validated by experience, may be used to determine the domain of competency.

A general framework based on these ideas is described below, and is complemented with analysis of illustrations by several examples.

General Framework

Character reasoning (CR) (Samsonovich & Aha, 2015) involves the concepts of a character and a character arc (these terms are explained below). Here characters are distinguished from actors. A **character** in CR is an abstraction, which is a virtual rational agent with its own goals, motives, senses, affordances, knowledge, and recent history (Haven, 2007, 2014). A **character type** is a class of characters given by a subset of character attributes (e.g., can be given by motives only). The top goal of a character may change in the course of character evolution, or **character arc**. This notion of a character is also distinct from a **role** in multi-agent planning literature (Campbell & Wu, 2011), which implies a fixed pattern of behavior. A **character arc** is the sequence of goals, intentions and other internal states experienced by the character through the narrative (Samsonovich & Aha, 2015). One actor can perform multiple characters, and vice versa. Selecting the types of characters and assigning characters to actors is called **casting**, or **characterization**. In team challenges, casting usually predetermines a solution of the problem (examples are presented below).

Character reasoning in AI belongs to the domain of narrative reasoning (Abell, 2009; Schmid, 2010; Finlayson & Corman, 2013), including narrative planning (Riedl & Young, 2010), which is different from other forms of planning in that all intentions and actions of actors in narrative planning must be motivated. In the present work, **believable character reasoning** (BCR) is understood as a kind of character reasoning in which goals, intentions and actions of an actor are justified by human-like motives. This rule applies to self and to other actors. Instead of providing a general criterion or definition for "human-likeness", it is assume that a list of human-like motives is given, and all other possible motives are considered not human-like and therefore not useful for BCR. Examples of human-like motives include:

- caring about a friend;
- caring about the team;
- caring about own safety;
- interest in an opinion or position that differs from the general consensus;
- tendency to rely on inductive reasoning that cannot be precisely explained verbally (intuitive reasoning);
- associating values with episodic memories.

To illustrate how BCR works, consider the following example scenario.

Example Scenario 1: Rescue Mission

The following proposed for consideration scenario conforms to the Freytag's (1863) narrative arc (http://www.ohio.edu/people/hartleyg/ref/fiction/freytag.html) with its standard components.

Exposition: An ad hoc rescue team with one limited radio communication channel has a mission to rescue a group of survivors from an area of a chemical explosion disaster, involving three buildings (A, B, C). The available information suggests that survivors are trapped in Building A, while the other two buildings are presumably empty.

One of the team members nominates himself and is elected to be the Leader. He will coordinate the team, others will follow his orders. The Leader orders the team to proceed to Building A.

Inciting incident: A phone call is received by the team via the radio channel from one of the survivors, who asks for help and tells that a group of people is trapped in a basement, but they cannot identify the building. The description does not match well Building A, although nothing can be concluded at this point. The phone dies. The source of the call cannot be verified.

Rising Action: The Leader does not have a good reason to change the plan, and decides to continue with it. However, one team member, Curious and Doubtful (CND), suggests exploring Buildings B and C. He gets Leader's permission to do this and proceeds to Building B. He finds that Building B does not have a basement, and Building C looks identical to B. At this point, CND decides to abandon his idea and to join the team. He communicates his decision to the team. Immediately after that, the team leader broadcasts a message that no survivors were found in the basement of Building A, and orders the entire team to explore upper floors of Building A. Then CND decides that he should go back to explore Building C. He tries to obtain permission of the Leader to do this, but the Leader orders him to go to Building A.

Climax: CND disobeys. He explores Building C and finds an entry to the basement obstructed by rubble. Sounds can be heard from the basement, but their nature cannot be identified. Based on intuitive reasoning, CND comes to believing that people are trapped in this basement.

Falling action: CND cannot remove the rubble by himself, and needs help of the rest of the team. He briefly describes the situation to the Leader and asks for help, but instead receives a reminder of the order to come to Building A.

Resolution: CND rebels. He takes an initiative and broadcasts a message with his own order to the team: to help him enter the basement of Building C, where he believes the survivors are trapped. He states that his decision is based on his interpretation of available to him information than cannot be transmitted by radio (visual, audio), and takes full responsibility for the consequences.

Dénouement: Several team members arrive at Building C to help CND. Together they remove the rubble, rescue survivors and give them necessary help. Then the rest of the team arrives. CND is elected as the new team leader. The team brings survivors to the base.

Let us now briefly analyze this example narrative. Three types of characters are involved in it:

- a potential team leader (one);
- a curious and doubtful opposition (one);
- a regular team member (several).

Selecting these types at the beginning may be justified by the initial settings, and is likely to predetermine the development of the narrative. Selecting other types may result in an alternative narrative, with an alternative outcome. These character types cannot be viewed as roles or goals in the sense in which these terms are used in the planning literature (Campbell & Wu, 2010). They are also distinct from actors: in principle, robotic actors performing these characters can be all identical to each other. Their differentiation is introduced at the time of characterization, or casting.

The main point of the above example is the believability of the protagonist character, CND, that proves useful. CND is believable, because he decides to do what he does being guided by the kinds of human-like motives listed above. Complementing his logical reasoning, these motives work as top-level heuristics in his decision making. Without the CND character, the basement of Building C may have not been found.

Therefore, this example serves as a qualitative argument in support of BCR. Before considering a numerical proof by example of the potential advantage of BCR, a general formalism should be outlined.

Formalism

Similarly to character reasoning introduced earlier (Samsonovich & Aha, 2015), the formalism of BCR is based on a top structure called a **hierarchical narrative network** (HNN), which is related to the notion of a narrative network (Pentland & Feldman, 2007), and is defined as the tuple:

$$\text{HNN} = \langle S,\ E,\ C,\ \mathcal{A},\ P \rangle, \qquad \text{(Eq. 1)}$$

where S is a set of nodes, E is a set of directed edges, C is a set of characters, $\mathcal{A}$ is a set of character arcs, and P is a set of performing actors. These terms can be explained intuitively as follows. An HNN includes a graph with a set of nodes S and a set of directed edges E. Here nodes represent actual and possible states (see the definition of a state below in this section) and fragments (i.e., fragments of the graph), and edges represent causal and temporal relations among nodes, inducing a partial order. Possible states, relations and rules of dynamics are given by the **domain theory**: a knowledge base that is assumed given, yet is not explicitly included in (Eq. 1). The network is hierarchical, because some of its nodes represent fragments of the same network. Fragments can be collapsed into nodes, and nodes can be expanded into fragments, as necessary.

In addition to S and E, an HNN (Eq. 1) includes a set of characters $C=\{c_j\}$, a set of possible character arcs $\mathcal{A}=\{A_j\}$, and the set of performing actors $P = \{P_i\}$. Our intuitive notion of a believable character was introduced above. Technically, a character c is represented by a tuple

$$c = \langle\ p,\ m,\ A\ \rangle, \qquad \text{(Eq. 2)}$$

in which the character is given by the perspective p, motives m, and the arc A. These terms need to be explained. The perspective p represents the current character's viewpoint, including senses of "now" and "here", "self" and "others" (own identity), and other contextual variables determined by embodiment (i.e., the performing actor P_i). The set of believable character's motives m includes human-like drives, values, and top-level guidance, that determine the selection of goals and intentions and usually do not change their nature within a character arc. A character arc A is a set of character's attitudes $\{a_i\}$, such as beliefs, goals, intentions, memories, percepts and affordances, taken at various moments of time. In general, a character c's attitudes are formed from states by attributing them to the character together with a certain modifier, e.g. (Samsonovich & Aha, 2015):

```
c.does(s0),
c.intends(s1),
c.ignored(s2),
c.achieved(s3),
c.saw(s4),
c.committed(s5).
```

From the definition (Eq. 1), one can observe that HNN combines a set of possible character arcs. As a result, a character in an HNN may not be committed to a particular arc or actor, and therefore may not have a unique perspective. Therefore, p and A in (Eq. 2) may remain unspecified in HNN. In general, a number of narratives can be derived from a given HNN. A **narrative** is obtained from HNN by resolving ambiguities and inconsistencies and assigning all characters to actors, all actions to intentions, and intentions to motives. As a result, each character c in a given narrative is committed to an arc A and to an actor p (the embodiment of c).

The notion of a narrative involves two important representations (Schmid, 2010): a fabula and a sjuzet. In the spirit of Abell's formalism (2009; 2011), a **fabula** is defined here as any part of an HNN written as the tuple (Eq. 1) that is internally closed (i.e., all intentions and actions are **motivated** and placed into arcs), consistent, and includes exactly one arc per character. Here "motivated" means that character intentions can be explained by, or derived from character motives. Also, a mapping of performing actors to characters needs to be specified in the fabula. Then, the **sjuzet** can be defined as the character arc corresponding to the storyteller character (usually the protagonist or the author; in our case of interest it is the character of the reasoning actor).

$$\text{Fabula} = \langle S^*, E^*, C^*, \mathcal{A}^*, P^* \rangle, \qquad \text{(Eq. 3)}$$

$$\text{Sjuzet: } c^* = \langle p^*, m^*, A^* \rangle, \qquad \text{(Eq. 4)}$$

The asterisks in (Eq. 3) refer to the working narrative subset of HNN, and in (Eq. 4) to the reasoner character. Definition of a state that is used here was given by (Samsonovich & Aha, 2015). A state $s \in S$ is defined to be a class of possible physical states of the world, such that a particular given fact applies to all those and only those physical states of the world. For example, a state can corresponds to a specified place and/or a moment or an interval in time, and/or can represent a particular object that is present there, or an event, a condition, a feature or property, etc., and any collection of them. States are therefore not necessarily mutually exclusive: for example, there may be more than one actual current state represented in the actor's working memory. Fractions and unions of states are themselves states. Therefore, states can be added and subtracted as sets. Controlled actions and processes are also states. Due to the hierarchical nature of HNNs, HNN fragments are also states, when represented by nodes. When a state is included in a narrative, it allows characters to form attitudes based on this state.

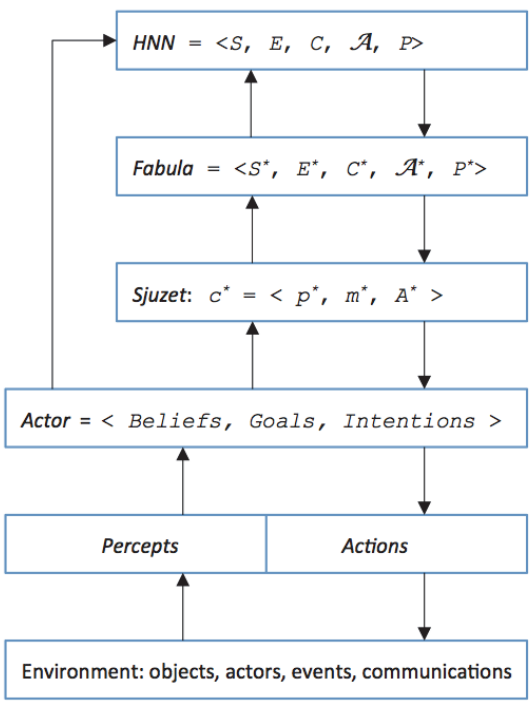

Figure 1. General architecture of a Believable Character Reasoner (see also Eqs. 1-4 and Algorithm 1).

Finally, a **Believable Character Reasoner** (Figure 1) is understood here as a system that implements the above formalism (Eqs. 1-4) and operates on an HNN, producing characters, character arcs, and a narrative, while being guided by human-like motivations. This means that goals, intentions and actions of a believable character are derived from human-like motives, drives, rules and heuristics, and other human-like top-level guidance. Accordingly, the actor performing a believable character given by the following tuple

$$\text{Actor} = \langle \text{Beliefs}, \text{Goals}, \text{Intentions} \rangle \qquad \text{(Eq. 5)}$$

should have goals (or desires) and intentions derived from the believable character.

The resultant general architecture of a Believable Character Reasoner is summarized in Figure 1, and the associated abridged top-level algorithm is given below. The structures of *HNN*, *Fabula*, *Sjuzet*, and *Actor* are given by (Eqs. 1-5).

Algorithm 1: Believable character reasoning (BCR) cycle

BCR (*Percepts*, *HNN*, *Fabula*, *Sjuzet, Actor, Actions*)
Input: *Percepts*, *HNN*, *Fabula*, *Sjuzet, Actor*
Output: *HNN*, *Fabula*, *Sjuzet, Actor, Actions*
Repeat
  Percepts ← perceive_input (Environment, Actor)
  [Actor, Sjuzet] ← orient (Actor, Sjuzet, Percepts)
  Parallel Thread 1:
    HNN ← top_update (Actor, Sjuzet, HNN)
    Fabula ← update_fabula (HNN, Fabula)
    Sjuzet ← update_sjuzet (Fabula, Sjuzet)
  Parallel Thread 2:
    Actor ← update_actor (Sjuzet, Actor)
    Actions ← produce_output (Environment, Actor)
  End
End

Thus, according to Algorithm 1, author-level BCR occurs continuously and indefinitely at all times of the operation of the actor, in parallel with the actor-level reasoning and decision making. However, the Fabula and Sjuzet only get updated under special circumstances (not in every cycle), when a new event or condition triggers substantial changes in HNN. This condition corresponds to a notable event in GDA (Klenk et al., 2013).

Numerical Test for the Concept

Consider the following example of a virtual world paradigm that illustrates the principles described above.

Example Scenario 2: Nuclear Disaster

A rescue team wearing protective suits and communicating via radio has a mission to retrieve expensive equipment from a nuclear disaster area ~500 meters wide, which is by now deserted. At the beginning, the team has a meeting in the center of the area (the cross in Figure 2). $N = 8$ actors are divided into pairs and assigned to each other and to locations on a circle, then proceed to their locations (Figure 2). Each actor works in close cooperation with the partner.

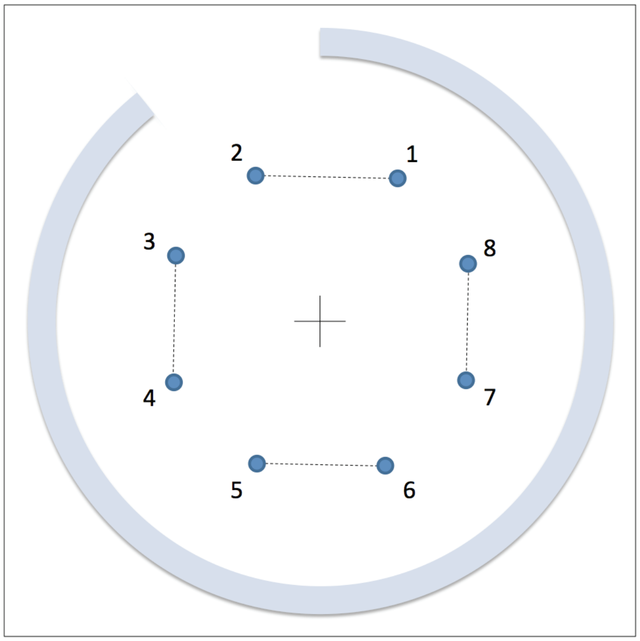

Figure 2. Positions of actors in the nuclear disaster scenario. Inaccessible area is shown by the grey circle with an opening.

When the actors are distributed around the circle as shown in Figure 2, a powerful nuclear explosion occurs, turning the area of previous disaster into complete ruins. Tall buildings around the perimeter collapse, closing most exits. To simplify further consideration, here it is assumed that nobody of the team died; however, some team members are injured and cannot move (assume that those who can move are not injured). The radio is lost. The range of visibility is less than 5 meters. In these conditions, the previous mission is obviously abandoned. The challenge for the team is to find and, if necessary, rescue each other, making sure that as many as possible (ideally, all) of the team members will be taken to safety, with minimal impact of radiation on their health.

The team knows the area and can navigate in it; however, they do not know which parts of the area became inaccessible (it is assumed that the interior of the circle in Figure 2 remains accessible). Therefore, one task for the team is to discover a path to safety by exploring the perimeter. In the configuration shown in Figure 2, only the exit to the North-North-West remains accessible, but the team cannot know this without exploration. Another task is to check all last known locations of the team members (or otherwise meet those who were located there) to make sure that necessary care is taken of all the injured. From the previous radio communications, the team knows each others' locations at the time of explosion, but cannot track further movements, or check immediately who is injured.

Non-BCR Approach

The new mission involves achieving the following goals:

(a) find a path to safety by exploring the perimeter;
(b) check the last known locations of missing team members;
(c) move the injured (and eventually self) to safety;
(d) share available information with others, if necessary, assigning tasks and roles to them.

While doing these tasks, actors will (unexpectedly or deliberately) meet each other. However, the probability of this event is assumed negligible, unless the actors are heading exactly toward each other or to the same known location. It is assumed that when two actors meet, they share all available information. In this case, they can also divide tasks among themselves, and decide to meet again at a certain location for further coordination. It is difficult to say at the beginning how this possibility can or should be used. It appears that there are four useful roles, or behaviors, corresponding to the above goals, that can be performed by individual actors with or without coordination with others:

(a) exploring the perimeter in search for an accessible exit to safety;
(b) checking the last known locations of all those actors who are presumably missing – in order to find the injured who are not taken care of;
(c) moving the injured to safety (assume that in order to move an injured team member outside of the area, the actor needs to know an accessible exit);
(d) coordinating others by serving as an information hub (e.g., this could be a useful role for an injured actor).

Assume that each actor creates a narrative network, which determines a set of possible narratives, based on currently available information and on a reasonable set of characters putatively assigned to team members, including self. This model will then be used to evaluate alternative narratives, e.g., based on simulations, and to determine the behavior that gives best expectations over many randomized trials.

In a non-BCR approach, an actor may take behaviors (a)-(d) as a basis, and determine own behavior as a choice among (a)-(d) with alternation probabilities taken as functions of time and new information, optimized through simulations. In these simulations, the actor will use the same model for other team members as for self, assuming that other team members reason similarly, except that their injured status may be different.

BCR Approach

Alternatively, the actor may start by reasoning as follows (this reasoning can be done automatically, e.g., using STRIPS (Fikes & Nilsson, 1971), if the corresponding actions and their effects are defined). Here the assumption is that this is done by the actor.

In order to be efficient, the team needs to exchange available information within itself. In order to exchange information in zero visibility without radio, the team needs to come together at one place. The best chance to come together in one place without prior coordination is to select the most salient place, which in this case would be the place of the team's previous meeting (the center), if it is still accessible. Therefore, the goal for each actor should be to come to the center to meet others. Then the team will determine who is missing and what the next goal will be.

The higher efficiency of this narrative compared to the previous one can be confirmed by simulations (as will be shown below), allowing the actor to select it as the working narrative. There are, however, two questions. Why would the rules used in this reasoning be available to the actor? And why should the actor assume that this sort of reasoning, and not the previous one, will be used by other team members in a heterogeneous team?

Both answers could be based on the believability estimate. Rules like "*in order to be efficient, the team needs to exchange available information within itself*" will be entered because they are part of the human common sense (and in this sense are believable). Moreover, general human-like motives listed in the previous Section will result in the formulation of the following goals for a healthy agent (thereby making the agent self-confident in its working narrative, according to our definition):

(e) Check the status of the partner (because of the worry about the partner).

(f) Meet with the team in the center (because of the worry about the team and the recent episodic memory of initial gathering in the center).

(g) Do not leave an injured partner alone (because of the worry about the partner).

At this point some existing goals are justified, while the actor may need to augment the narrative with new goals, adding one step before going to the center: check the partner's location. Then, if the partner is healthy, he is likely to do the same, and the two are likely to meet half-way to each other. If the partner is injured, he will be found and taken to the center. If the partner is missing at the location, then he will be presumed healthy. Therefore, this step will not compromise efficiency of the narrative, and may be helpful.

Numerical Test

Which method, BCR or non-BCR, is more efficient? The question seems nontrivial, and the answer can be found through numerical experiments. To do this, the scenario described above was simulated in Matlab using a BCR strategy (with meeting of the team at the center after the explosion) and a non-BCR strategy (when healthy actors practically have no chance to meet each other). The motion of agents was simulated using graphical animation (Figure 3). The results are the following. In 44 runs, the time spent by the team in the radioactive area is significantly shorter in the case of a BCR strategy, as compared to the non-BCR strategy (t-test p-value: 1.7e-23, Wilcoxon rank sum test p-value < 6.8e-16, histograms are shown in Figure 3).

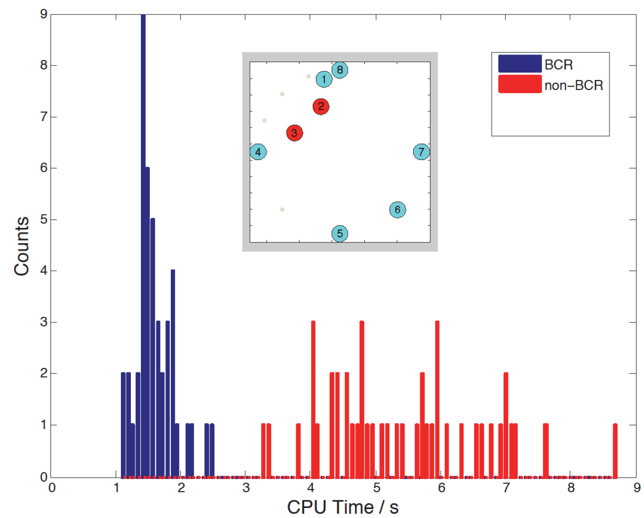

Figure 3. Histograms of the time spent by the team in the disaster area before all team members are taken to safety. Blue: BCR strategy, red: non-BCR strategy. The inset shows a typical snapshot during the session. Actors 2 and 3 are injured, the rest are together exploring the perimeter after they met at the center.

Discussion

Traditionally, the notion of self-confidence of an autonomous artificial intelligent agent includes three aspects: competence in the domain of assignment, adequacy of available information, and the ability to quantify and report the first two. This paper takes a different approach, assessing the concept of believability as one possible measure of self-confidence. Here the key notions can be finally redefined as follows. A narrative involving multiple characters is believable, if it is interpretable in terms of human psychology and, in particular, human theory-of-mind attributed to the characters. An actor is believable, if (s)he or it is guided by a believable narrative. A character is believable, if its behavior is consistent with human-like motives in the context of a believable narrative. An actor is **self-**confident, if it evaluates its knowledge of the current situation and the relevant general knowledge as reliable.

At first, it may seem hard to accept a connection between these notions of a believable character and a self-confident character. Indeed, suppose that an auto mechanic is called on to perform a surgery, while he lacks any medical background or skills. It seems that in this situation, a believable character must be uncertain in its own knowledge and decisions, and therefore not self-confident. What is missed by this logic, however, is that a character in this situation can be self-confident (in the above sense) in its understanding of the situation and other actors involved in it, as well as in their understanding of him. This could make a huge difference. For example, if I know that surgery is in principle doable and is the only chance to save my partner's life, and there is nobody else around to help, then I am confident that I must attempt the surgery, even if I do not have skills and prior experience. I am also confident that the partner believes the same, and will cooperate (if the partner is believable), even if I look clueless and uncertain. This gives me a chance to succeed.

In this sense, the link between self-confidence and narrative reasoning exists for a particularly narrow, though useful, form of self-confidence. Essentially, the agent is confident in its assumptions about other agents when it can be reasonably confident in its own and others' believability. This allows the agent to reason and act predictably and to expect the same from others (believable behavior is more predictable).

Linking believable behavior to human-like behavior also seems problematic, because there is no unique, well-defined, predictable human-like behavior in any given situation. Not all humans behave similarly; in fact, they frequently differ radically and unpredictably from each other. Nevertheless, they have the ability to understand even those who are different. The differences in behavior can be accounted by the differences in individual motives, values, reasoning schemes, etc. This does not mean that all those behaviors cannot fit into one low-dimensional subspace of human-like behaviors, as opposed to any algorithm-driven or random behavior.

In general, the agent's confidence in his beliefs about other agents (and, by extension, plans based on those beliefs) will be greater to the degree that the other agents behave like the agent would expect them to behave based on its model of self. This is the leverage of agent believability in a team.

To summarize, this work presented a general approach to producing and measuring believable character reasoning in autonomous actors working in a team. Believable behavior is more predictable, which gives an advantage to the team in unexpected challenging situations. Believable behavior is defined here as interpretable based on human-like motives, which makes BCR-actors human-compatible, creating further benefits for heterogeneous teams.

The simple computational illustration used here indicates that believability can increase the efficiency and robustness of the team as a whole, as compared to more traditional approaches in non-narrative (Campbell & Wu, 2010) and narrative planning (Riedl et al., 2008; Riedl & Young, 2010; Ware & Young, 2014). While the considered example may not be generally conclusive, it suggests that the idea deserves further investigation. This will be done elsewhere.

In the presented scenario of a nuclear disaster, all actors were performing one and the same character type, which simplified the consideration. However, in a more general consideration, one could assume that at the moment of the narrative-changing event, each actor is performing one of several character types in the team, which (due to the believability requirement) will predetermine his new character choice in the new situation (even though the actors may be physically identical). E.g., the Leader may take responsibility to instruct all team members he meets to go to the center. This may further improve the robustness of solutions found by actors, because other team members will know a priori what new character role each of them is likely to take.

Speaking more generally, biological inspirations in goal reasoning and autonomy (Samsonovich, 2014), in this case introduced via believability, can be beneficial for teams operating in uncertain situations, and in this sense have a general value for AI.

Acknowledgments

The author's many thanks go to Dr. David W. Aha from the Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory (Code 5514), Washington, DC. The funding for this research was provided by OSD ASD (R&E). The views and opinions contained in this paper are those of the author, and should not be interpreted as representing the official views or policies, either expressed or implied, of OSD or NRL.

References

Abell, P. (2009) A Case for cases: Comparative narratives in sociological explanation. *Sociological Methods and Research*, 38(1):38-70.

Abell, P. (2011). Singular mechanisms and Bayesian narratives. In: Demeulenaere, Pierre, (ed.) *Analytical Sociology and Social Mechanisms*, pp. 121-135. Cambridge University Press, Cambridge, UK. ISBN 9780521154352.

Campbell, A., & Wu, A.S. (2010). Multi-agent role allocation: Issues, approaches, and multiple perspectives. *Autonomous Agents and Multi-Agent Systems*, 22: 317-355. DOI: 10.1007/s10458-010-9127-4.

Fikes, R.E. & Nilsson, N.J. (Winter 1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence,* 2 (3–4): 189–208. doi:10.1016/0004-3702(71)90010-5.

Finlayson, M. A., & Corman, S. R. (2013). The Military Interest in Narrative. *Sprache und Datenverarbeitung*, 37 (1-2).

Freytag, G. (1863). Die Technik des Dramas. Published by S. Hirzel. https://archive.org/details/dietechnikdesdr01freygoog

Haven, K. (2014). *Story Smart: Using the Science of Story to Persuade, Influence, Inspire, and Teach*. Santa Barbara, CA: ABC-CLIO, LLC. ISBN: 9781610698115.

Haven, K. (2007). *Story Proof: The Science Behind the Startling Power of Story.* Westport, Connecticut: Libraries Unlimited. ISBN 978-1-59158-546-6.

Klenk, M., Molineaux, M., & Aha, D. W. (2013). Goal-driven autonomy for responding to unexpected events in strategy simulations. *Computational Intelligence*, 29(2), 187-206. doi: 10.1111/j.1467-8640.2012.00445.x

Pentland, B. T., & Feldman, M. S. (2007). Narrative networks: Patterns of technology and organization. *Organization Science*, 18(5), 781-795. doi: 10.1287/orsc.1070.0283

Riedl, M.O., Stern, A., Dini, D., and Alderman, J. (2008). Dynamic experience management in virtual worlds for entertainment, education, and training. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning,* 3(1).

Riedl, M.O. & Young, R.M. (2010). Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research* 39: 217-268.

Roberts, M., Vattam, S., Aha, D.W., Wilson, M., Apker, T., & Auslander, B. (2014). Iterative goal refinement for robotics. In: A. Finzi & A. Orlandini (Eds.) *Planning and Robotics: Papers from the ICAPS Workshop.* Portsmouth, NH: AAAI Press.

Roberts, M., Vattam, S., Alford, R., Auslander, Apker, T., Johnson, B., & Aha, D.W. (2015). Goal reasoning to coordinate robotic teams for disaster relief. In A. Finzi, F. Ingrand, & Andrea Orlandini(Eds.) *Planning and Robotics: Papers from the ICAPS Workshop.* Jerusalem, Israel: AAAI Press.

Samsonovich, A. V. (2014). Goal reasoning as a general form of metacognition in BICA. Biologically Inspired Cognitive Architectures, 9: 105-122. DOI: 10.1016/j.bica.2014.07.003.

Samsonovich, A. V. and Aha, D. W. (2015). Character-oriented narrative goal reasoning in autonomous actors. In: Aha, D. W. (Ed.). Goal Reasoning: Papers from the ACS Workshop. Technical Report GT-IRIM-CR-2015-001, pp. 166-181. Atlanta, GA: Georgia Institute of Technology, Institute for Robotics and Intelligent Machines. https://smartech.gatech.edu/bitstream/handle/1853/53646/Technical%20Report%20GT-IRIM-CR-2015-001.pdf#page=169

Schmid, W. (2010). *Narratology: An Introduction*. Walter de Gruyter GmbH & Co. KG, Berlin/New York. ISBN 978-3-11-022631-7.

Ware, S.G., and Young, R.M. (2014). Glaive: A state-space narrative planner supporting intentionality and conflict, in *Proceedings of the 10th Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE2014)*, pp. 80-86. Raleigh, NC.