The IDEA of Us: An Identity-Aware Architecture for Autonomous Systems

CARLOS GAVIDIA-CALDERON, The Alan Turing Institute, London, United Kingdom and The Open University, Milton Keynes, United Kingdom
ANASTASIA KORDONI, Lancaster University, Lancaster, United Kingdom
AMEL BENNACEUR, The Open University, Milton Keynes, United Kingdom
MARK LEVINE, Lancaster University, Lancaster, United Kingdom
BASHAR NUSEIBEH, The Open University, Milton Keynes, United Kingdom and Lero - University of Limerick, Limerick, Ireland

Autonomous systems, such as drones and rescue robots, are increasingly used during emergencies. They deliver services and provide situational awareness that facilitate emergency management and response. To do so, they need to interact and cooperate with humans in their environment. Human behaviour is uncertain and complex, so it can be difficult to reason about it formally. In this paper, we propose IDEA: an adaptive software architecture that enables cooperation between humans and autonomous systems, by leveraging in the social identity approach. This approach establishes that group membership drives human behaviour. Identity and group membership are crucial during emergencies, as they influence cooperation among survivors. IDEA systems infer the social identity of surrounding humans, thereby establishing their group membership. By reasoning about groups, we limit the number of cooperation strategies the system needs to explore. IDEA systems select a strategy from the equilibrium analysis of game-theoretic models, that represent interactions between group members and the IDEA system. We demonstrate our approach using a search-and-rescue scenario, in which an IDEA rescue robot optimises evacuation by collaborating with survivors. Using an empirically validated agent-based model, we show that the deployment of the IDEA system can reduce median evacuation time by 13.6%.

CCS Concepts: • **Human-centered computing → Collaborative interaction**; • **Computer systems organization → Robotic autonomy**.

Additional Key Words and Phrases: autonomous systems, game theory, social identity, agent-based modelling

1 INTRODUCTION

> I'm so aware of you, waiting for me. But I won't realise this, until you do. This idea of us.
>
> Jono McCleery, This Idea of Us

Mass emergencies are events or situations with serious consequences for life and property, that require special arrangements and coordination between multiple emergency response agencies [22]. For example, terrorist attacks on critical infrastructure, earthquakes, and fires in large buildings qualify as mass emergencies. These

Authors' addresses: Carlos Gavidia-Calderon, The Alan Turing Institute, London, United Kingdom and The Open University, Milton Keynes, United Kingdom, cgavidia-calderon@turing.ac.uk; Anastasia Kordoni, Lancaster University, Lancaster, United Kingdom, a.kordoni@lancaster.ac.uk; Amel Bennaceur, The Open University, Milton Keynes, United Kingdom, amel.bennaceur@open.ac.uk; Mark Levine, Lancaster University, Lancaster, United Kingdom, mark.levine@lancaster.ac.uk; Bashar Nuseibeh, The Open University, Milton Keynes, United Kingdom and Lero - University of Limerick, Limerick, Ireland, bashar.nuseibeh@open.ac.uk.

ACM 1049-331X/2024/3-ART
https://doi.org/10.1145/3654439

incidents are inherently disruptive: they negatively affect the operation of systems, as they require a quick response to minimise their impact. When they occur, *first-responders* —like firefighters, paramedics, and police officers— coordinate and perform complex activities under time and physical constraints to assist the people affected [96]. First-responder teams are increasingly relying on autonomous systems —like rescue robots and drones— to support their effort [11].

The affected populations also play an active role during mass emergencies, especially during the evacuation of fellow survivors in the aftermath of the emergency. Some survivors, also called *zero-responders*, immediately mobilise to render assistance to those in need [27]. Zero-responders contribution is crucial to saving lives and provide assistance before professional first responders arrive. In major incidents, zero-responders can easily reach difficult locations and provide urgent care [78]. Social psychology research has established that, in the aftermath of a major incident, most zero responders help fellow survivors to evacuate the dangerous area. We refer the reader to the excellent paper by Drury for evidence of this phenomenon [27]. Even after first response teams arrive, most zero responders keep supporting the rescue effort, by either working together with first-responders or coordinating the evacuation among themselves. Zero-responder support is crucial when first responders are limited in number, or operating in a different part of the affected area [3]. Whilst cooperation between zero and first responder is critical for saving human lives, accounting for help allocation in a synergistic and coordinated manner has been problematic during mass emergencies [97]. In this paper, we show that autonomous systems can promote cooperation between first and zero responders, by modelling and reasoning about survivors' group behaviour.

Modelling group behaviour, and human behaviour in general, is challenging due to the uncertainty of what humans do (or do not). Planning for human-system cooperation requires accounting for the needs, expectations, and preferences of humans [71, 151]. Many psychological theories explain, specify and represent human behaviour based on a set of behaviour determinants [71]. However, only a few focus on the principles that determine group behaviour, and even less are applied to emergency settings [105]. The *social identity approach* [120] focuses on social structures and group membership to explain human behaviour. It states that, alongside personal identity, humans also have multiple identities that derive from the social groups to which they belong. Humans tend to behave according to the values and expectations imposed by their social groups, when these social identities become salient [124].

During emergencies, most zero-responders develop a *shared identity*, as a consequence of experiencing danger together [28]. Sharing a social identity favours pro-social behaviours, such assisting other survivors during emergencies [27]. A minority of zero-responders will not develop the shared identity, and may still prioritise their own safety, favouring a pro-self behaviour [135]. We build upon the social identity approach to enable cooperation between autonomous systems and the humans in their environment. In scenarios where social identity is the main driver of human behaviour, as in emergencies [18], autonomous systems can reason in terms of groups instead of individuals. Our proposal, called IDEA, is an adaptive software architecture where autonomous systems infer the social identity of the humans they interact with, and use this information to compute the optimal strategy for cooperation.

Identity inference happens under uncertainty: even for humans, it is hard to know the social identity of other people. In this paper, we use a game-theoretic approach to deal with the uncertainty of identity inference, when computing the optimal strategy for cooperation. *Game theory* is a modelling framework that supports decision-making under uncertain agent preferences [79]. It is extensively used to model human behaviour during emergencies [50, 62, 63, 149, 150], interactions between humans and autonomous systems [74, 82, 136, 148], and social identity dynamics [5, 133]. In IDEA, autonomous systems build *games of incomplete information* of their interactions with humans. Using game theory, we obtain a predicted behaviour for the human and an optimal strategy for the autonomous system. The autonomous system's strategy is the best response to the predicted human behaviour, and vice versa. There is empirical evidence that, when model assumptions are met, humans adopt the behaviours predicted by game theoretic models [93, 137, 142].

To evaluate IDEA, we built upon the IMPACT model developed by van der Wal *et al.* [131]. The IMPACT model is an agent-based simulation of emergency evacuations that has been validated against data from evacuation drills. This model is widely used to study the role of trained staff [37] and communication strategies [130] during emergency evacuations. To alleviate first-responder workload —and reduce evacuation time— we designed an IDEA rescue robot that, for evacuating a victim unable to move, will request help from survivors in the area. To maximise the chances of getting support, the robot targets zero-responders who adopt a shared identity. We show that deploying the IDEA rescue robot reduces evacuation time and that its impact depends on the severity of the emergency.

This paper makes the following contributions:

- We represent interactions between autonomous systems and humans in their environment through a game-theoretic model based on social identity principles.
- We develop an end-to-end adaptive software architecture for identity-aware autonomous systems, called IDEA.
- We build upon an empirically-validated agent-based model of emergency evacuations to show that deploying IDEA rescue robots reduces the evacuation time of survivors.

We made the agent-based model, the IDEA agent, and the data gathered public [46], hoping it will be of use for researchers interested in human-autonomous system cooperation.

This paper extends our previous work [42]. Our short paper outlined a software architecture for adapting to uncertain human preferences, and explored its effectiveness using synthetic data. This work improves it in four ways. First, we developed IDEA: a refinement of our original architecture, specialized on adaptation based on group membership and social identities. Second, we developed IMPACT+: an extension of an agent-based model of emergency evacuations that supports the deployment of IDEA agents. Third, we used IMPACT+ to evaluate the impact of IDEA rescue robots on evacuation time, and investigate the factors that affect its performance. And fourth, in this paper we provide an extended discussion of social identity, game theory, and their application to emergency contexts.

2 A SEARCH-AND-RESCUE EXAMPLE

First-responder teams are frequently under stress during emergency evacuations [97]. The number of victims, the criticality of their injuries, and a challenging terrain, represent obstacles to the evacuation effort [67]. Fortunately, they are not the only ones supporting victim evacuation: often other survivors in the disaster zone also engage with evacuating survivors. Contrary to popular belief, zero responders do show solidarity when facing adversity, due to the emergence of a shared identity among people experiencing the same emergency situation [27]. Emergency response demands optimising scarce resources, including time and personnel. Professional first-responder teams can greatly benefit from the solidarity that emerges among survivors. Survivors, by helping some victims, can ease the workload of the first-responder team, that can devote their efforts to other people needing help.

Autonomous systems, like drones, are increasingly deployed in disaster zones, mainly to provide situational awareness [33]. However, autonomous systems can also support and coordinate the efforts of both first-responders and zero-responders during emergency evacuations [132]. Cooperation scenarios between autonomous systems and humans may fall into two categories: 1) *control scenarios*, where the autonomous system is viewed as a tool commanded by a human, and 2) *teamwork scenarios*, where humans and autonomous systems jointly work towards accomplishing a common goal [99]. Unlike control scenarios, teamwork scenarios require socially-aware autonomous systems to succeed [55]. In this paper, we focus on the teamwork category and refer the interested reader to related work for the first category [64]. As such, the autonomous system will be an active actor during

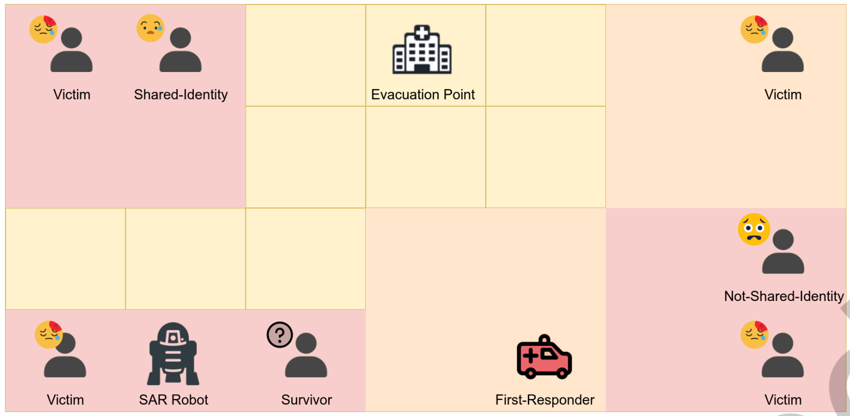

Fig. 1. A Search and Rescue example. The rescue robot traverses the emergency zone, looking for survivors for evacuation. When the victim is not able to move, the robot needs help from surrounding survivors. In some situations, like in the top-left corner, the survivor has developed a shared identity, so it will be willing to help the victim. In some others, like in the bottom-right corner, the survivor has *not* developed the shared identity, so it is better to contact the first-responder team for victim evacuation.

the evacuation effort —along with first-responders and other survivors— working towards the goal of evacuating the disaster zone in the shortest time.

More specifically, let us consider a small search-and-rescue (SAR) robot —like the Hector robot [89]— in charge of locating survivors during an emergency. The rescue robot in Figure 1 is capable of autonomous exploration of the affected area and survivor recognition. When finding a survivor that is able to move, the rescue robot can offer guidance to the evacuation point using its speakers. However, when the victim found is unable to move, the rescue robot needs help for their evacuation, due to its small size. In these situations, the robot can request support from the first-responder team. However, first responders are scarce and mostly busy in most major incidents. The rescue robot can also request support from other survivors in the surrounding area. It is expected that some survivors will be willing to help the rescue robot with emergency evacuations, but others might prioritise their personal safety.

In this paper we propose IDEA: an adaptive architecture that builds on the social identity approach and game theory to engineer a collaborative rescue robot. An IDEA rescue robot uses game theory to model cooperation scenarios with survivors and first-responders (section 4). In these models, IDEA robots calculate the likelihood of receiving help from survivors using social identity markers (section 5). By optimising the participation of survivors and first-responders in the rescue effort, IDEA robots can reduce evacuation time (section 7). Prior to presenting IDEA, in section 3 we introduce the social identity approach and game theory.

3 PRELIMINARIES

In this section, we describe the social identity approach, its importance in emergency situations, and how we can use game theoretic models to reason about it.

3.1 The Social Identity Approach

The social identity approach is a social psychological approach that explains collective behaviour at mass gatherings [58] and evacuations [29]. This approach comprises the *social identity theory* [125] and the *self-categorisation theory* [128]. According to the social identity theory, alongside a personal identity, people also have multiple social identities that derive from the social groups to which they belong. People sharing the same social identity, i.e., belonging to the same group, tend to behave more favourably towards members of the same group [120]. Research on crowd behaviour has shown that, when a crowd develops a sense of belonging to the same group, they become members of a psychological crowd [104]. The psychological crowd is different from a physical crowd of individuals, who happen to be in the same space without a sense that they are joined in a meaningful group [141].

Self-categorisation theory refers to the process where a person categorises themselves as a group member, at a given time. Ascribed social categories, such as gender and race, are forms of social identity that provide a basis for self-definition [26]. There are multiple social categories, as there are multiple social groups. For example, profession and family status are different social categories, and being a software engineer or a parent are their associated social identities. People can be in multiple categories and belong to different social groups. Which social identity, and thereof social category, will be activated depends on the given context. For example, a software engineer who experiences a flood where they live, will behave in accordance with the identity of the survivor towards their neighbours (fellow survivors), but as a software engineer in a software engineering conference. This happens through a process of depersonalisation, a cognitive transformation through which individuals tend to see others in their group as more similar to themselves [120].

Collective behaviour occurs when people shift from their personal identity to their group identity, endorsing norms and behaviours associated with this group. Acting upon a group identity can produce pro-social behaviours towards other members of the same group [77]. This social identity adoption is expressed via *identity markers*, and they are diverse in nature. For example, they can be associated with stable social categories, such as gender, age or cultural background [38]. They are also related to how close people are in space, they way people walk, or how they speak [70].

Previous work showed that natural language can be indicative of group membership. Certain linguistic categories can be used as identity markers for a shared identity [30]. Using words and expressions that reflect one's affiliation with a group are indicative of the adoption of a shared identity [87]. For example, using the pronoun "we" increases the salience of a group identity [98], and develops a sense of community [129]. Also, using high inclusivity language is associated with strong group membership ties [52].

The Search-and-Rescue Example. Social identity plays a role in decision-making during emergencies. The social identity model of collective psychosocial resilience establishes that, during an emergency, the sense of common fate favours the emergence of a *shared identity* among survivors [27, 144]. In an emergency context, a crowd of strangers is turned into a psychological crowd with common goals [78]. This means that individuals who may be holding conflicting identities outside the emergency context, have these differences overridden by the common experience of threat leading to the development of a shared identity [94]. Survivors sharing an identity provide support to each other, expect to be supported, and cooperate towards common goals [78]. However, not every survivor develops a shared identity. There is evidence of individualistic, or pro-self, behaviour in some emergency evacuations [39]. These survivors act in their own interest, without cooperating or coordinating with others.

3.2 Game Theory

Game Theory studies scenarios, called *games*, where rational self-interested agents interact, and these interactions have an impact on the payoff agents receive. Examples of such games include board games, card games, markets, and even software development practices. For example, a match in a chess tournament is a game, having chess players as agents and the points per match as the payoff agents try to maximise. During a game, agents interact via *actions*. In chess, actions are sequential and correspond to, for example, moving a piece to a specific board position. An agent's *strategy* determines the actions they adopt during a game. In a software engineering context, we can model bug prioritisation using game-theory [48]. Bug reporters are the agents, and their payoff is a function of how many of their reported bugs are fixed. Their actions in the model are the priority of the bugs they report, and strategies depend on how accurate these priorities are (either inflating priorities or deflating them).

There are numerous outcomes in a game like chess, having each player adopting a specific strategy. Under its rationality assumptions, game theory predicts that only a subset of these outcomes, called *Nash equilibria*, is possible. At Nash equilibria, an agent's strategy is the best response to the strategy of their opponent. These outcomes are stable, since a deviation from the equilibrium strategy would result in a payoff loss. In some domains, like professional sports and internet auctions, there is empirical evidence that humans adopt equilibrium strategies when engaging in games [93, 137, 142].

In the chess match example, game state and agent preferences are transparent and visible to both agents. In contrast, in a poker match this is not the case. Agents have partial visibility of the game state, given that the deck and opponent hands are hidden. This applies to other situations, such as salary negotiations, where the parties involved hide their preferences. Or in sealed-bid auctions, where a bidder's real valuation of the auctioned item is not public [101]. To model this uncertainty, *games of incomplete information* map agent preferences to *types*. Types represent information private to agents, like cards in hand in poker, or a candidate's compensation expectations when negotiating salary. Assuming a common-knowledge probability over these types, we can obtain the Nash equilibrium of a game of incomplete information. At these outcomes, called *Bayes-Nash equilibria*, an agent's strategy defines actions for every potential type assignment.

The Search-And-Rescue Example. There is a large body of research on using game theory for modelling human behaviour during emergency evacuations [50, 62, 63, 149, 150]. Evacuations fit the game definition: survivors constitute the agents, their behaviours are the actions, and their payoff can be expressed —for example— as evacuation time. Under those modelling considerations, as in real evacuation scenarios, an agent's payoff is not only a function of individual behaviour, but depends on the actions of the rest of the agents.

Game theory allows the prediction of psychological propensities and their associated behaviours, during human-system interactions. These predictions can inform an autonomous system's adaptation strategy. In the following section, we present an approach for engineering autonomous systems capable of inferring identity and promoting cooperation, using game-theoretic models based on social identity principles.

4 MODELLING COOPERATION BETWEEN HUMANS AND AUTONOMOUS SYSTEMS

In this section, we start with an overview of IDEA (*"Identity-Aware Adaptive Architecture"*). Then, we describe the game-theoretic models of human-system interaction used within IDEA systems.

4.1 An Overview of IDEA

In this work, we propose IDEA: a software architecture for *self-adaptive* autonomous systems, that change behaviour according to the perceived social identity of the humans in their environment. IDEA systems need to reason about the humans they interact with, the identities they might act upon, and how these identities impact the system's goals. Game theory is well-suited for reasoning about the interactions between people sharing a social identity [5], as their interactions have a direct impact on the payoff they perceive. Social identities have norms and values: when a member follows these norms, other members benefit from the affirmation of their identity [120]. In contrast, members violating these norms threaten the identity of themselves *and* the complying members, producing anxiety and discomfort. Consequently, compliant members can choose to penalise offending members [133]. Using game theory, we can model members of a social group as agents, their behaviours as actions, and their payoff functions as a function of compliance to identity norms. Economists have successfully used this modelling framework for explaining identity-driven behaviours that escape classic economic analysis [5].

During operation, an IDEA system follows the steps shown in Figure 2. First, it estimates the identity of the person it is interacting with. The *identity estimator* component calculates a probability distribution over a set of

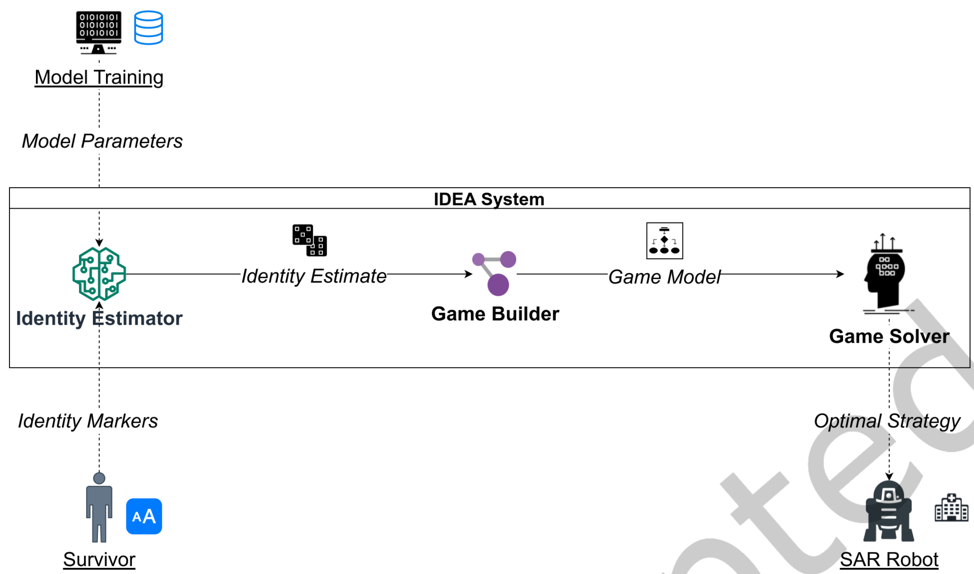

Fig. 2. An overview of IDEA systems: Based on the identity markers of the people they interact with, the identity estimator calculates a probability for each candidate identity. These values are used by the game builder component to model the human-system interaction using game theory. Finally, the game solver uses the model to compute a strategy, for the IDEA system to enact.

candidate identities. Then, the *game builder* component uses these probability values to build a game model of the human-system interaction. The *game model* includes candidate actions for the IDEA system, and their payoff given human reactions. Finally, the *game solver* component uses this model to compute the optimal strategy for the IDEA system, ensuring it maximises its payoff.

To implement this functionality, we adopt the MAPE-K reference framework for self-adaptive systems [68]. The framework establishes that the autonomic manager of a self-adaptive system is composed of four elements —*Monitor*, *Analyse*, *Plan*, and *Execute*— that operate over a shared *Knowledge* element. In IDEA, the Knowledge element is composed by game-theoretic models. In these models, discussed in detail in subsection 4.2, we represent interactions between autonomous systems and the humans in its environment. Within the model, the system's actions are determined by its actuators capabilities, and its payoff function represents the systems' mission. Human actions, candidate identities, and payoff functions are based on social psychology research. IDEA systems manage a set of these game-theoretic models. System designers should tailor each model to specific identity-driven adaptation scenarios. The model catalogue is defined at design time, while some model parameters are obtained at runtime by IDEA components, like game builder and identity estimator.

The Search-And-Rescue Example. A rescue robot adopting IDEA should build game-theoretic models of its interactions with emergency survivors. The model has two agents: the rescue robot and a survivor. The robot, via its actuators, requests survivor assistance to evacuate a victim. The payoff the survivor receives from helping others depends on their adoption of the shared identity. If they adopt the shared identity, they will be more prone to help.

Detecting a survivor's adoption of the shared identity is not straightforward: arguably, this is challenging even for humans [56]. An IDEA system perceives the world via its sensors, and calculates probabilities of identity adoption based on the data sensed. For example, in previous work, microphones would capture what survivors say during the emergency [70]. In this case, identity estimation is based in the presence or absence of linguistic features indicative of shared identity adoption, as informed by psycholinguistic evidence (e.g., "we" versus "I").

We can model the uncertainty in shared identity adoption using games of incomplete information. In our scenario, the IDEA rescue robot builds a game of incomplete information each time it needs assistance from a survivor. This model has two candidate types for the survivor: 1) adopting a shared identity, and 2) *not* adopting a shared identity. Games of incomplete information require a probability distribution over these types, that the robot infers from sensor data. The robot's actuators determine its actions in the game-theoretic model. Survivors react to these actions, obtaining a payoff defined by shared identity adoption. Once the model is ready, the IDEA rescue robot computes its Bayes-Nash equilibria. This computation produces, per equilibrium, a strategy for the survivor and a strategy for the rescue robot. The rescue robot's strategy defines its adaptation approach, that is the best response to the predicted behaviour of the survivor. In the next subsection, we describe this game-theoretic model in detail.

4.2 Models of Human-Autonomous System Interaction

IDEA systems use games of incomplete information to model interactions between the autonomous system and humans in its environment. Such models, represented by game trees like the one in Figure 3, have the following elements:

- A set of two agents: the autonomous system S and a human H.
- A set N_E of terminal nodes, where the game ends.
- A set A_S of actions for the autonomous system, and the set A_H of actions for the human.
- A set N_S of choice nodes, where the autonomous system can perform an action. And the set N_H of choice nodes for the human.
- A set I of candidate identities for the human. Each candidate identity $i \in I$ is associated to a node $n \in N_S$. These nodes have the same available system actions.
- A probability distribution $P : I \mapsto [0, 1]$ over candidate identities.
- Two payoff functions, $u_S : N_E \mapsto \mathbb{R}$ for the autonomous system and $u_H : N_E \times I \mapsto \mathbb{R}$ for the human.
- At an autonomous system's choice node $n \in N_S$, the action function $\chi_S : N_S \mapsto 2^{A_S}$ determines which actions are available for the autonomous system to perform.
- There is also an analogous action function $\chi_H : N_H \times I \mapsto 2^{A_H}$ for the human agent.
- The successor function for the autonomous systems $\sigma_S : N_S \times A_S \mapsto N_H \cup N_E$ selects the node to follow after the system performs an action.
- The human also has a successor function $\sigma_H : N_H \times A_H \times I \mapsto N_S \cup N_E$, that selects the node after the human performs an action.

These models are *games of incomplete information in extensive form, with chance moves* [79], with some additional restrictions. These restrictions include considering only two agents, only supporting types (candidate identities) for the human agent, and modelling a single human-system interaction. These game parameters have a large impact on the size of the resulting model. Without these conditions, the IDEA system might produce intractable models, or take a long time for devising an adaptation strategy. Large models can be specially challenging for autonomous systems with limited computing resources.

The Search-and-Rescue Example. In previous work, we proposed autonomous systems that use survivor expressions to calculate the probability of shared identity adoption [70]. In the example, the rescue robot incorporates this probability into a game-theoretic model of its interaction with a survivor. The equilibrium analysis of this model will determine if it is better to request help from them or to contact the first-responder team.

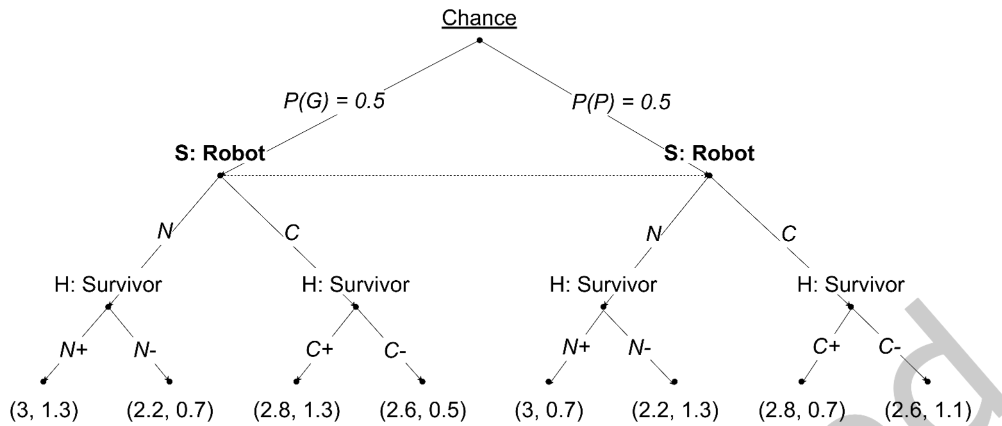

Fig. 3. Game-theoretic model of the search-and-rescue scenario: The survivor has two candidate identities, adopting a shared identity (G) or not (P). The game starts with the robot either asking from survivor help (N), or contacting the first-responder team (C). After that, the survivor can either: help with victim evacuation (N_+); follow the robot leaving the victim behind (N_-); wait for the first-responder team to arrive (C_+); or leave the scene by themselves (C_-).

Table 1. Elements of our game-theoretic models of human-system interaction, along with their values for the SAR example

Symbol	**Description**	**SAR Example**
S	Autonomous system agent.	SAR robot.
H	Human agent.	Survivor.
A_s	System actions.	C: Request first-responder support. N: Request support from a survivor.
A_H	Human actions.	C_+: Wait with victim until help arrive. C_-: Evacuate without the victim. N_+: Follow the robot with the victim. N_-: Follow the robot without the victim.
I	Candidate identities.	G: Adopts the shared identity. N: Does not adopt the shared identity.
u_S	System's payoff function.	Expected number of successful evacuations.
u_H	Human's payoff function.	Score based on probability of evacuation and identity adoption.

We will focus on one identity-driven adaptation scenario: the rescue robot finding a victim unable to move, and another survivor in close proximity. The elements for this model are listed and described in Table 1. There, the robot has two possible actions in $A_S = \{C, N\}$. These actions are: 1) to either call for first-responder support (C), or 2) request support from the survivor and navigate with the victim to safety (N).

It is in the best interest of the emergency response effort to maximise the number of evacuations supported by survivors. This would alleviate first-responder workload, allowing them to support the evacuation of more victims. The rescue robot is part of the emergency team, so its payoff function u_S is the expected number of successful evacuations. A survivor that evacuates following the robot, or assisted by a first-responder or other survivor, has a probability of 1.0 for evacuating successfully. If a survivor is able to move and evacuates by themselves, this probability is 0.8. For a victim without support and unable to move, this probability is 0.2.

In the game-theoretic model, the survivor reacts to the robot's action. The survivor actions $A_H = \{C_+, C_- N_+, N_-\}$ can either agree with the robot's proposal or disagree with it. When the robot offers first-responder support (C), agreement translates to waiting with the victim until help arrive (C_+). Disagreement with this proposal implies evacuating by themselves (C_-), leaving the victim with the robot. If the robot requests for help for victim evacuation (N), agreement requires for the survivor to follow the robot (N_+), carrying the victim with them. Disagreement translates to the survivor following the robot, leaving the victim behind (N_-).

Regarding the survivor's payoff function u_H, we assume a base payoff tied to their probability of successful evacuation (1.0 for assisted evacuation, and 0.8 for unassisted evacuation). As mentioned before, the survivor's identity has an impact on the payoff they perceive. In our game-theoretic model, the survivor agent has two candidate identities $I = \{G, P\}$. Survivors adopting a shared identity (G) are expected to support the group and assist with victim evacuation. If they behave according to these expectations, they increase their payoff in 0.3. Otherwise, they decrease their payoff in 0.3. When survivors *do not* adopt the shared identity (P), they receive the 0.3 bonus when they leave the victim behind, prioritising their personal safety. If they do assist with victim evacuation, their payoff is diminished in 0.3.

Figure 3 represents this game model. For example, in the bottom-left node, the robot requests help (N) from a survivor adopting the shared identity G, and they accept the request, taking the victim with them (N_+). In this case, the robot's payoff is $u_S = 3$, reflecting the evacuation of 3 people: the survivor, the victim, and another survivor rescued by first-responders. For the survivor, their payoff is $u_H = 1.3$. This value is the sum of 1 for a successful evacuation plus 0.3 for following identity expectations.

Payoff functions, survivor responses, and candidate identities must be based on empirical evidence. We build on existing literature and our understanding of the domain to set these model parameters for the motivating example.

5 IDEA: AN IDENTITY-AWARE ARCHITECTURE FOR AUTONOMOUS SYSTEMS

In Figure 4, we show the five software components that constitute an IDEA system. These components fulfil three main responsibilities: 1) using identity-related sensor readings X, they build a game tree G of the human-system interaction, 2) they estimate the human's identity adoption probability $\hat{P}$ and incorporate it to the model G, and 3) use the model G to decide a strategy s_S for the system. In this section, we describe how these responsibilities are accomplished by the software components within IDEA.

5.1 From Sensors to Game-Theoretic Models

Not every human-robot interaction requires an identity-driven adaptation. For example, the search-and-rescue scenario we described in section 2 does not require the rescue robot to adapt when encountering professional first-responders. Its goal is to leverage survivor's solidarity via adaptation. The role of the Game Selector component —from the *Monitor* element— is to detect identity-driven adaptation scenarios during the system's operation.

The Game Selector component gathers data from the autonomous system's sensors to determine the need for an identity-driven adaptation. It also provides information to the Game Builder component to complete the game-theoretic model associated to that adaptation. In Figure 4, we observe that the *Game Selector* component notifies the *Game Builder* component that the identity-driven scenario G' is taking place, with sensor readings X associated to this interaction.

The Search-And-Rescue Example. In section 4.2, the game-theoretic model has two runtime parameters: 1) $P(G)$, the probability that the survivor in the proximity adopts the shared identity and 2) $P(P)$, the probability that they do not. We set values for all the other game parameters at design time. As explained in subsection 3.1, the

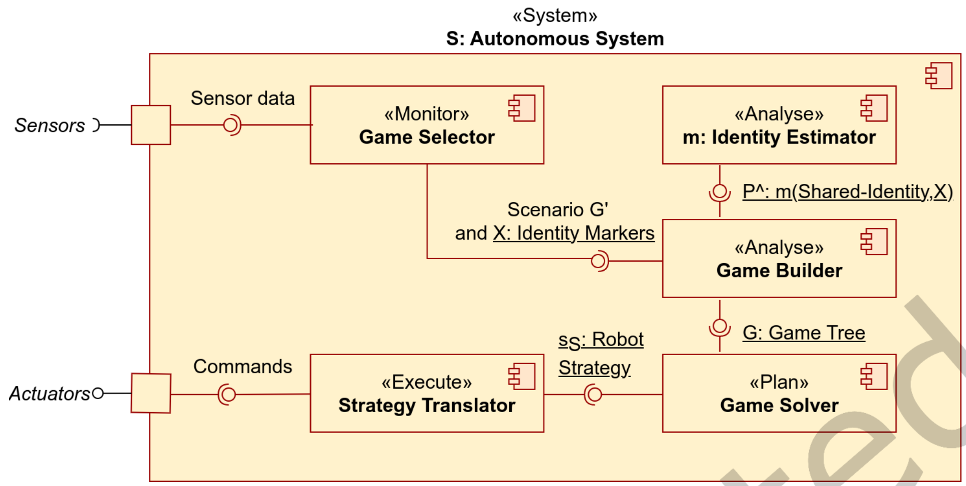

Fig. 4. Component diagram of IDEA: the game selector triggers an identity-driven scenario G', sending identity marker information X to the game builder. This software component assembles the game tree G, incorporating the identity probabilities $\hat{P}$ produced by the identity estimator. The game solver calculates the robot strategy s_S, that the strategy translator transforms into actionable actuator commands.

IDEA robot calculates identity adoption probabilities from survivor's identity markers. In the example, these markers are oral expressions. The responsibility of the Game Selector is to use sensors —like a microphone array— to capture the survivor's expressions and convert them to text for further processing at the Identity Estimator component.

5.2 Estimating Identity Adoption

Arguably, the most important model runtime parameter is the probability distribution over candidate identities P. In IDEA, the *Identity Estimator* component has the responsibility of calculating $\hat{P}$ —an estimate of P— based on sensor readings X.

The Game Selector component uses the system's sensors to look for identity markers. When found, it forwards the sensor readings X associated to the detected markers. The Identity Estimator internally uses a classifier $m_\theta : I \times \mathbb{R}^n \mapsto [0, 1]$ to compute probability values for each identity $i \in I$ given a vector X of n sensor readings. In Figure 4, the *Game Builder* component produces a game-theoretic model G, like the one in Figure 3, from a given identity-driven scenario G' and sensor readings X, integrating the output of the Identity Estimator component into the resulting model.

The Search-and-Rescue Example. In our previous work [70], we developed an Identity Estimator for identifying survivor expressions that suggest the adoption of a shared identity. Labelling a survivor expression as indicative of shared identity can be seen as a text-classification problem. For this reason, we fine-tuned a pretrained BERT model [60, 127], over a dataset of expressions we extracted from videos of survivor accounts. After training, our Identity Estimator obtained a Brier score[1] of 0.12, suggesting its predictions have acceptable quality. For further details, please refer to the code and data available at GitHub [46].

Besides identity estimation based on verbal expressions, there are other markers we could use during emergencies and other contexts. For example, Philpot and Levine performed spatial clustering in running and walking

[1]The Brier Score measures the error in probability predictions. It varies between 0 and 1, with lower values suggesting better performance [8].

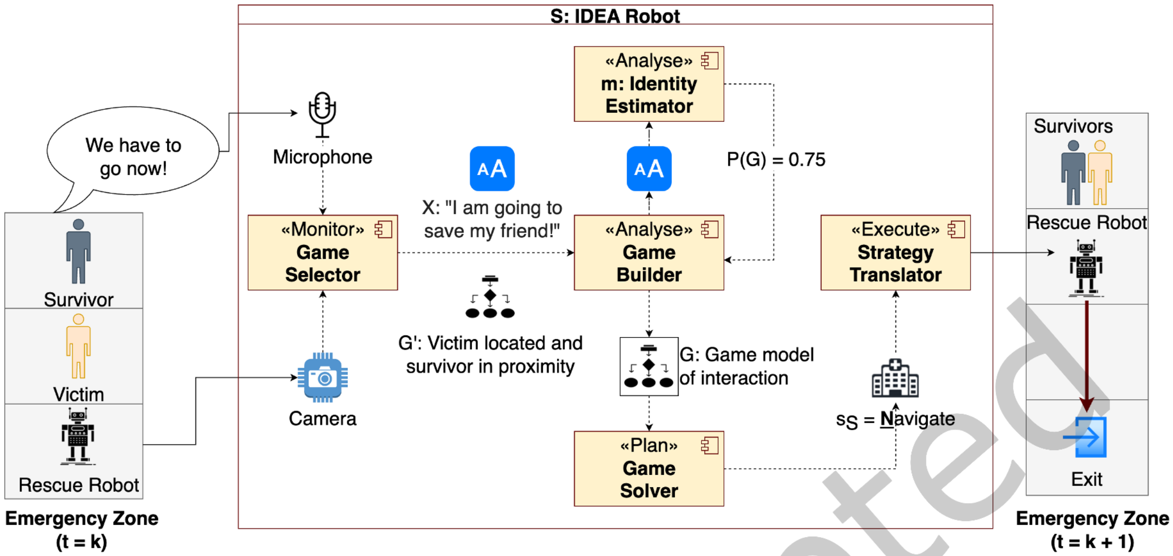

Fig. 5. An interaction within the search-and-rescue scenario. The IDEA rescue robot detects a victim, a survivor in their proximity, and a survivor expression. The Game Selector component identifies this situation as the identity-driven scenario G', with the expression as the survivor's identity marker X. The Game Builder builds the game tree G of this interaction, including the probability of shared identity adoption ($P(G) = 0.75$). The Game Solver component finds that, for the model G, the robot action should be to guide the victim and survivor to the evacuation point ($s_S = N$).

behaviour, using digital visual data of a train evacuation [94]. Similarly, Van der Wal *et al.* estimated identity based on speed and distance in which a social interaction is still possible [131]. Identity was also estimated based on the mirroring effects of mental state between people (imitating emotional face expressions) [107, 145].

Once the Identity Estimator is trained, the IDEA rescue robot can use it for identity inference during operation. When engaging with a survivor, the Identity Estimator receives the survivor's expressions as input, to output the probability that they share an identity with a victim (Figure 5). The Game Builder component incorporates this value into the game-theoretic model of the identity-driven scenario (section 4.2), and forward the updated model to the Game Solver.

5.3 Calculating the System's Strategy

The Game Builder component produces a game-theoretic model G of the interaction of the IDEA system with a human. The role of the *Game Solver* component is to calculate the Nash equilibria of G, using one of the several algorithms available for this purpose [92]. This software component produces two strategies per equilibria $s*$: 1) one strategy s_H for the human agent and 2) another strategy s_S for the autonomous system. As discussed in subsection 3.2, the equilibrium strategy for the autonomous system s_S is the best response to the predicted strategy s_H for the human agent.

The Game Solver component can produce more than one equilibrium, where each system strategy s_S assigns a probability to each system action in $\chi_S(n), n \in N_S$ in the game-theoretic model G. The role of the *Strategy Translator* component is to translate these system strategies into executable actuator commands. Transforming strategies to actuator commands can be straightforward in some scenarios. For example, there is a direct mapping between actions and commands when obtaining a single equilibrium, with a single action with probability 1.0.

In more complex scenarios, engineers need to design a heuristic to select the actuator commands to execute, given the equilibria obtained. The design goals of this heuristic are context-dependent, and can be driven by risk minimisation, optimising response time, or any other metric relevant to the system's context.

The Search-And-Rescue Example. The Game Solver component produces a strategy s_H for the survivor, and the corresponding best response strategy s_S for the IDEA rescue robot. If the equilibrium strategy for the IDEA robot is to request first-responder support ($s_S = C$), the Strategy Translator component would obtain the robot's position from its localisation system, and send it to the emergency team. In case the equilibrium strategy requires requesting help from the survivor ($s_S = N$), the Strategy Translator would notify the survivor (e.g., via the speakers) and, if the survivor agrees, the IDEA robot would use the navigation system to guide the survivor to the victim's location.

6 EVALUATION

As is common practice in the self-adaptive systems domain, we used an illustrative example —emergency evacuation— and a simulation model to evaluate IDEA's performance [90]. Due to the nature of emergencies, using simulation allows us exploring multiple configurations and scenarios safely, ethically, and in a cost-effective way [53, 99, 131]. While many simulation techniques have been used to model emergency evacuations, *agent-based models (ABM)* dominate the space in recent years. Unlike techniques like systems dynamics or queuing networks, ABMs represent multiple autonomous decision-making agents reacting to the environment, like in real-world crowd evacuations [53].

ABMs of emergency evacuations differ in the nature of the social interaction they represent. For example, Zia and Ferscha model social influence mediated by technology [150]. Christensen and Sasaki focus on the interaction of disabled victims with the environment and terrain [21]. Singh and Padgham model survivors taking detours for evacuating with their loved ones [117]. While all these models represent real and relevant behaviours during emergencies, they do not explicitly model the emergent solidarity due to shared identity adoption. Researchers like von Sivers *et al.* have included this phenomenon in their simulation model [135], however its code it is not freely available.

In contrast, the IMPACT model, developed by van der Wal *et al.* [131], is open-source and does represent social identity solidarity. The IMPACT model is an agent-based simulation of the evacuation of a transport hub, and it has been validated against data from evacuation drills. Among others, researchers have used IMPACT to study the role of trained staff [37] and communication strategies [130] during emergency evacuations. To examine IDEA systems in situations closer to real rescue scenarios, we extended IMPACT to support IDEA robot agents. We call our IMPACT extension IMPACT+, and its code is available online [44]. In subsection 6.1, we describe IMPACT+ in detail.

6.1 IMPACT+: An Evacuation Simulation for IDEA agents

The IMPACT model is an agent-based simulation of an emergency evacuation, developed with the NetLogo modelling environment [143]. In this model, survivor behaviour is affected by social factors. For example, survivors help victims, groups are formed and move together, and behaviours spread [38, 131] via social contagion[2]. We based IMPACT+ on the version of the model developed by Formolo *et al.*, given that it incorporated staff support into the simulation [37]. Figure 6 shows a frame of IMPACT+[3].

The IDEA *Agent.* The original IMPACT model has agents representing survivors and staff. For IMPACT+, we developed an additional agent, representing an IDEA rescue robot. This agent moves randomly during the

[2]Social contagion refers to the process of dissemination of emotions, intentions, and beliefs within a group [13].
[3]Click here for an animation of the agent-based model in action.

Table 2. Comparison table between the IMPACT model and IMPACT+, our extension for evaluating IDEA systems.

Concept	IMPACT	IMPACT+
Agents	Survivors and staff	Survivors, staff, and IDEA rescue robot
Identity markers	Age, gender, and cultural cluster of survivor X_S and victim X_V.	Same markers as IMPACT.
Survivor support	During evacuation, survivors encounter victims and decide to stay with them based on their identity markers.	Survivors can approach victims after been invoked by the IDEA rescue robot.
Staff support	Staff members notify survivors about their closest exit.	Staff members can be invoked by the IDEA rescue robot to speed-up victim recovery.
Victim recovery	When accidents happen, a victim and supporting survivors do not move for a fixed amount of time.	The victim's recovery time is reduced when receiving help from first-responders f, other survivors s, and survivors with robot support s^r.

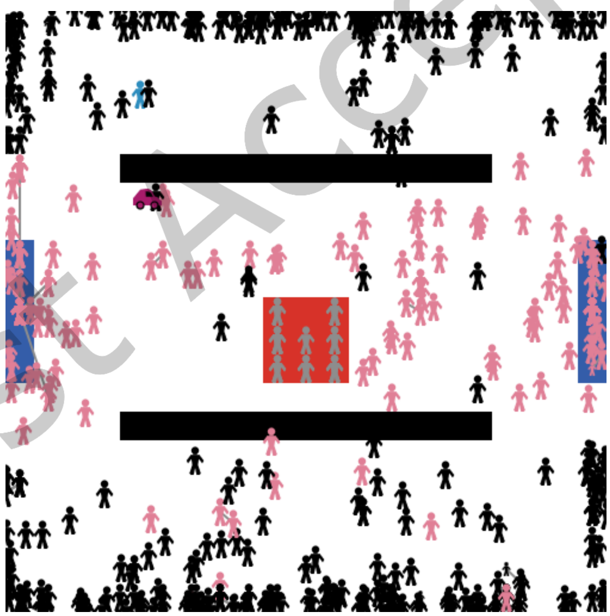

Fig. 6. Agent-based model of emergency evacuation [44]: the IDEA robot (in magenta) looks for victims (in orange). When the robot finds a victim, it can either request assistance from first-responders (in light blue) or from other survivors (pink if they are evacuating, black if they are not).

evacuation, looking for victims within its range. By victims, we mean survivors that cannot continue with the evacuation, due to events like injury or accident.

When locating a victim, the IDEA agent has two possible actions: 1) request help from another survivor in the area (N), or 2) request first-responder assistance (C). For contacting another survivor, the IDEA agent needs to approach their location to be within its range. Contacting a first-responder can be done remotely, so the IDEA agent does not need to move to make contact.

Victims. In the original IMPACT model, survivors can fall during the evacuation. Fallen survivors, along with other survivors helping them, do not move for a fixed amount of time f. After f time units pass, they recover and continue with the evacuation. The probability of falling depends on survivor speed and how crowded are their surroundings. Survivors moving fast in crowded areas are more likely to fall.

In IMPACT+, we refer to fallen survivors as *victims*. Also, fall length f is *not* fixed: it can be reduced when a victim receives support from other survivors or first-responders. The magnitude of the reduction is determined by the helping effect H_i of the agent $i \in \{f, s, s^r\}$ providing support. First-responder agents (H_f), survivors (H_s), and survivors with robot support (H_{s^r}) have different helping effect values. We consider that $H_s < H_{s^r} < H_f$.

Survivors. The original IMPACT model simulates passengers evacuating a transport hub. In this model, passenger evacuation depends on socio-cultural factors [131]. Evacuation speed, compliance to staff instructions, and helping other passengers, vary among the agent population, depending on age, gender and cultural cluster[4].

A passenger S leading an evacuation group can help a fallen passenger V within its range, according to a probability. This probability is a function of the age, gender, and the cultural cluster of *both* passengers S and V. Due to this, in IMPACT+ we use the age, gender and cultural cluster of both passengers (X_S and X_V) as the identity markers $X = \{X_S, X_V\}$ for inferring the adoption of a shared identity. These markers are well-suited for social identity estimation, as they remain stable over evacuation time. By contrast, verbal expressions of shared identity, as we proposed in section 2, are more dynamic in nature. Both static and dynamic identity markers can inform identity estimation and complement each other. However, IMPACT and IMPACT+ are agent-based models where agents do not communicate verbally, so we cannot use oral expressions as identity markers section 2.

IMPACT authors relied on research on the social identity approach to design agent behaviour [38]. Extensive empirical evidence show the identity markers within the model explain group behaviour in real-life emergencies [131] [141]. For example, research has shown that members of the same group are more likely to help each other [29], and that children are more likely to receive help [32]. These findings translate in IMPACT+ into concrete probability values. For example, the probability of a male adult passenger S helping an elderly female passenger V from the same cultural cluster is 40%.

In IMPACT+, *survivors* are the passengers evacuating the disaster zone. We modified their helping behaviour: instead of just staying with the victims until they recover —as in the original model— survivor support reduces the time the victim is not moving. The reduction in time is determined by the helping effect value of the survivor. Helping effect has a value of H_s for survivors and H_{s^r} when they have IDEA robot support. We configure H_s to be small. When a survivor helps a victim due to an IDEA robot request, H_{s^r} increases the effectiveness of the survivor's assistance, to reflect the IDEA robot's contribution to victim treatment [91].

First-responders. The original IMPACT model has two types of agents: passengers and *staff members*. Staff agents represent security professionals, with knowledge of the disaster area [37]. During evacuation, staff agents move randomly, notifying passengers about the most convenient exit for their evacuation. Depending on the staff member's training level, passengers can or cannot comply with their instructions.

We extended the staff member agent capabilities in IMPACT+, so they can fulfil *first-responder* duties. Like survivors, they have a helping effect H_f, so they can reduce the time victims are not able to move. The value

[4] *Cultural clusters* are groups of countries according to cultural similarities [108]

of H_f is larger than H_s and H_{sr}, to reflect first-responder training and experience. First-responders can answer requests for help from the IDEA agent. If they are available, they move to the position of the victim —notified by the IDEA agent— to start providing assistance.

6.2 An IDEA System for IMPACT+ Simulations

Figure 7 shows an IDEA system (section 5) that supports survivor evacuation within IMPACT+ (subsection 6.1). In this section, we describe its most important components.

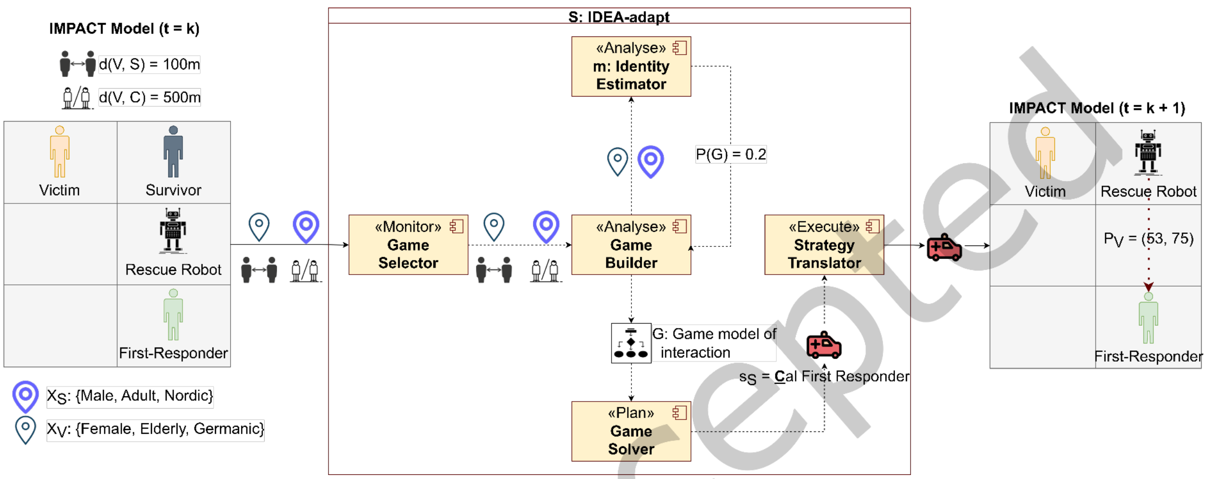

Fig. 7. An interaction during an IMPACT+ simulation run. When the robot agent locates a victim, the Game Selector receives the identity markers of the victim X_V and the closest survivor X_S, along with the victim's distance to this survivor $d(V, S) = 100$ and the closest first-responder $d(V, C) = 500$. The Game Builder produces the game tree G of this interaction, using the probability of shared identity adoption $P(G) = 0.2$ inferred from X_V and X_S. The Game Solver processes G and suggests contacting the first-responder team ($s_S = C$), communicating the victim's position $P_V = (53, 75)$.

From Sensors to Game-Theoretic Models. In the agent-based model, survivors have agency regarding helping victims: they choose either to help them or ignore them. In contrast, their interactions with first-responders are uni-directional: survivors only receive information from first-responders regarding the closest exit. To reflect this, the game theoretic model G —produced by the Game Builder component— does *not* include a survivor response when the robot requests first-responder support ($s_S = C$).

Within G, the payoff function for the survivor u_H outputs the following values: $u_H = 1$ when the survivor acts according to its identity, $u_H = -1$ when their behaviour *is not* aligned to their identity, and $u_H = 0$ when they do not interact with the IDEA agent. Preliminary experiments showed that performance of the IDEA agent is greatly affected when its payoff function u_S *does not* incorporate environmental information. Hence, we made u_S depend on both distance and first-responder availability. As seen in Figure 7, the Game Builder is aware of the victim's distance to the closest survivor $d(V, S)$ and to the closest first-responder available $d(V, C)$. This information is used within the Game Builder to calculate u_S. For example, when requesting help from first-responders, $u_S = -1$ if the first-responder is not available, $u_S = 1$ when the first-responder is available and closer than surrounding survivors, and $u_S = 0$ if the first-responder is available but further than other survivors in the area.

Estimating Identity Adoption. In subsection 6.1, we noted that within IMPACT+ simulations we use the gender, age, and cultural cluster of the survivor S and the victim V as the identity markers $X = \{X_S, X_V\}$. Figure 7 shows the Identity Estimator component m_θ receiving an encoded version of X as input, to output a probability estimate $\hat{P}(G) = m_\theta(G, X)$ of the survivor S helping the victim V due to a shared identity. We implemented m_θ as a multilayer perceptron (MLP), with a single output neuron with a sigmoid activation function.

Figure 8 shows the process of training m_θ. Initially, we gathered data of survivor-victim interactions by running IMPACT+ without deploying an IDEA agent. Per interaction, we recorded the identity markers —gender, age, and cultural cluster— of both the survivor (X_S) and the victim (X_V), as well as the outcome Y of the interaction. $Y = 1$ when the survivor helped the fallen victim, and $Y = 0$ otherwise. We further split the dataset in two: we used 77% of the data to estimate the parameters θ of candidate MLPs, and 33% for assessing model quality.

Within IMPACT+, and likely in real-world scenarios, the helping behaviour of survivors is probabilistic. This implies that, for example, a male adult survivor *will not always* help an elderly female victim within the simulation. This happens only in 40% of their interactions, and can be adjusted via IMPACT+ parameters. We need the probabilities produced by m_θ to be as accurate as possible, so we used the *Brier score* (BS) to evaluate its probabilistic predictions. BS measures the mean squared difference between the probability of shared adoption $\hat{P}(G)$ —produced by m_θ— and the actual outcome Y within the simulation. BS produces values between 0 and 1, with lower values indicating better predictions. Our final Identity Estimator had a Brier score of 0.25, which we deemed suitable for evaluation purposes.

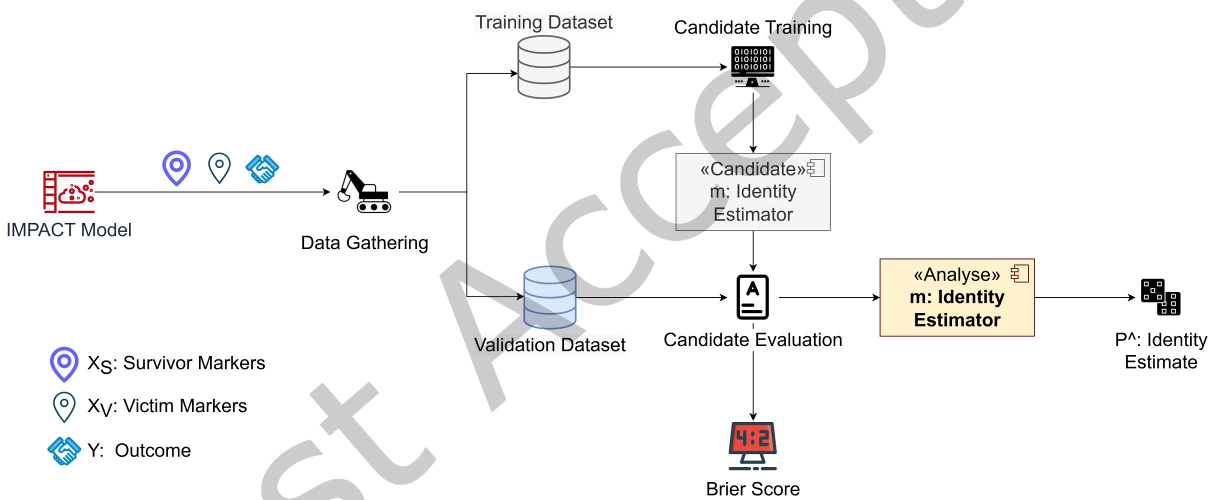

Fig. 8. Training the Identity Estimator for IMPACT+ simulations. We used IMPACT+(subsection 6.1) to gather data of survivor-impact interactions. Per interaction, we gather markers of the survivor (X_S), the victim (X_V), and the outcome (Y, 1 if the survivor helped the victim, 0 otherwise). We trained each identity estimator candidate over the training dataset, using the validation dataset to assess the quality of its predicted probabilities using the Brier Score. We use the best performing Identity Estimator (m_θ) for estimating shared identity adoption ($\hat{P}(G)$) during evaluation (subsection 7.1 and subsection 7.2).

Calculating the System's Strategy. We used the Gambit software tool [86] to calculate the Nash equilibria of the game-theoretic model of the robot-survivor interaction. The Game Solver obtains equilibria with two characteristics: 1) has pure strategies, and 2) are subgame perfect. *Pure strategies* require agents to select *one* of the actions they have available. For the IDEA robot, this means either request first-responder assistance ($s_S = C$) *or* request help from a survivor in close proximity ($s_S = N$). We adopt pure strategies for simplicity, to avoid dealing with the randomisation aspect of mixed strategies. By focussing on *subgame perfect equilibria*, we ensure the robot adapts to a rational response from the survivor. If we do not find equilibria fulfilling these two requirements, the IDEA agent calls for first-responder assistance to minimise risk.

Table 3. Input and output variables of the IMPACT+ agent-based model of emergency evacuations.

Input	Description
N_s	Number of survivors.
S_a	Number of survivors travelling alone.
S_g	Number of survivors travelling in groups.
S_c	Number of survivors from cultural cluster c.
N_f	Number of first-responders.
N_e	Number of exits.
C	Cultural clusters.
H_s	Helping effect for survivors.
H_{sr}	Helping effect for survivors with robot support.
H_f	Helping effect for first-responders.
f	Fall length for victims, in seconds.
Output	**Description**
t	Evacuation time, in seconds.

7 RESULTS

The purpose of IMPACT+ is to examine the effect of deploying IDEA systems in the emergency response effort. IMPACT+ is an extension of IMPACT, so it simulates major social categories along with their behaviours and norms. These characteristics made IMPACT+ a good testing bed for IDEA systems. Using a simulation model like IMPACT+ allows us to measure IDEA system's performance under multiple settings. Environmental factors, like the severity of the emergency, affect the spread of social identities and the decision-making of their members [97]. In this section, we explore these influences to better understand the performance of IDEA systems in dynamic emergency scenarios.

7.1 Evacuation Time Reduction

Evacuation time, defined as the time it takes for *all* survivors to leave a disaster zone, is a metric frequently used to evaluate emergency evacuation efforts [83, 117, 118]. In this section, we investigate if the adoption of IDEA systems reduces evacuation time.

We used IMPACT+ (subsection 6.1) to measure evacuation time. We compared the performance of three IDEA variants: 1) the *proself-oriented* variant $\text{IDEA}_{\text{self}}$ operates under the assumption that survivors will not help the victim, so it always requests first-responder support ($s_S = C$), 2) the *prosocial-oriented* variant $\text{IDEA}_{\text{social}}$ assumes survivors will always help surrounding victims, so it always requests help from other survivors ($s_S = N$), and 3) the $\text{IDEA}_{\text{adapt}}$ variant uses identity markers and game theoretic models to decide if $s_S = C$ or $s_S = N$. $\text{IDEA}_{\text{adapt}}$ is a proper implementation of our approach (section 5), while $\text{IDEA}_{\text{self}}$ and $\text{IDEA}_{\text{adapt}}$ are non-adaptive variants. For baseline purposes, we also measure evacuation time without robot support.

Simulation settings. Table 3 contains the parameters the IMPACT+ model. For evaluation purposes, we are using the default values from Formolo's *et al.* IMPACT implementation [37]. Hence, we have 800 survivors ($N_s = 800$) evacuating a $400m^2$ room with 2 exits ($N_e = 2$) and 3 sections. 50% of survivors are travelling alone ($S_a = 400$), while the rest are travelling in groups ($S_g = 400$). Survivors are diverse: they include men, women, children, and the elderly. Also, they are divided in 11 equally-sized cultural clusters ($|C| = 11, |C_i| \approx 73, i \in \{1, \ldots, 11\}$). There are 8 first-responders in the area ($N_f = 8$), able to attend requests for help.

The helping effect for survivors (H_s), first responders (H_f), and survivors with IDEA agent support (H_{s^r}) are new parameters introduced in IMPACT+. While the original IMPACT model has a constant value for fall length ($f = 30$), in our extension we can modify the value of f. We assume first responders can reduce recovery time by $H_f = 50\%$ per time unit. Survivors can also reduce victim recovery time, but only by $H_s = 2\%$ per time unit. When survivor assistance is mediated by an IDEA agent, recovery time reduces in $H_{s^r} = 22\%$. The values for H_s, H_f, and H_{s^r} remain constant over all simulation runs, and they reflect our understanding of the contribution of each agent to victim recovery. The most impactful is first-responder support, followed by survivors with the assistance provided by an IDEA agent.

Results. We explore if the adoption of IDEA has a positive impact in the evacuation effort, by reducing evacuation time t. To this end, we obtained t samples —via IMPACT+— in four configurations: 1) without rescue robot support, 2) with the support of IDEA$_{\text{adapt}}$, 3) with the support of IDEA$_{\text{self}}$, and 4) with the support of IDEA$_{\text{social}}$.

We posit that the effectiveness of our adaptive architecture varies with the magnitude of the disaster. Severe disasters require more resources for a timely evacuation, so the contribution of survivors becomes more impactful. To explore disasters of various severity levels, we simulated evacuations with 20 values of fall length $F = \{30, 60, 90, \ldots, 570, 600\}$. We consider that severe disasters generate serious accidents on survivors, that take more time to recover. Per each $f \in F$ value, we obtained a set T of evacuation time samples, where $|T| = 100$. For example, Figure 9 shows evacuation time samples for the 4 configurations when $f = 360$. We posted complete metrics for all scenarios at our GitHub repository [45].

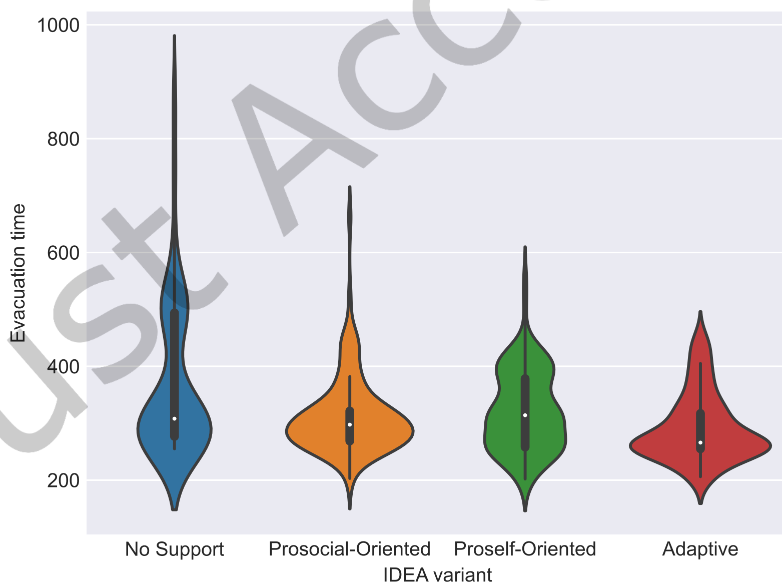

Fig. 9. Violin plot of evacuation time samples T per IDEA variant (IDEA$_{\text{social}}$ in orange, IDEA$_{\text{self}}$ in green, IDEA$_{\text{adapt}}$ in red, and no IDEA support in blue) when fall length $f = 360$. Per variant, we show the probability density of evacuation time, its median (white dot), and its interquartile range (black bar).

Table 4. Evacuation time median ($\tilde{T}$) per fall length value (f, in seconds). Per f, the configuration with minimum $\tilde{T}$ is shown in bold, and in italics configurations *not* significantly different.

f	No Robot	$\text{IDEA}_{\text{self}}$	$\text{IDEA}_{\text{social}}$	$\text{IDEA}_{\text{adapt}}$
30	309.0	*268.5*	292.5	**265.0**
60	314.0	*276.5*	*288.5*	**272.5**
90	310.5	**262.0**	296.5	*265.0*
120	314.0	**267.0**	290.0	*277.0*
150	310.0	**265.0**	295.5	**265.0**
180	310.0	**262.0**	282.0	*268.0*
210	308.0	*280.5*	*289.0*	**268.5**
240	309.5	*283.5*	*292.5*	**276.0**
270	309.0	*286.5*	*293.0*	**273.0**
300	301.5	*276.0*	*298.5*	**274.5**
330	314.0	*299.0*	*302.0*	**273.0**
360	308.0	314.0	297.5	**266.0**
390	302.5	*311.0*	*309.0*	**272.0**
420	315.0	*313.5*	**305.5**	*311.5*
450	320.0	**284.0**	*286.5*	*312.0*
480	311.0	**282.0**	*298.0*	*285.0*
510	309.5	*299.0*	*300.5*	**270.5**
540	*302.0*	*311.5*	**297.5**	*298.0*
570	315.0	**274.0**	*293.0*	**274.0**
600	310.0	*292.5*	*298.5*	**272.0**

Table 4 shows in bold the configuration with the minimum median evacuation time $\tilde{T}$ for each $f \in F$. We observe that $\text{IDEA}_{\text{social}}$ is the best performer —i.e. has the smallest $\tilde{T}$— only when fall length $f \in \{420, 540\}$. $\text{IDEA}_{\text{self}}$ has the smallest $\tilde{T}$ in 7 of the 20 fall length values in F. In the other 13 scenarios, $\text{IDEA}_{\text{adapt}}$ did best.

For each $f \in F$, we performed a *Kruskal-Wallis H test* [72] to verify there are statistically significant differences in the value of $\tilde{T}$ between the 4 configurations. The test indicates a significant difference ($p < 0.05$) for *every* $f \in F$: these results indicate there are differences in the value of $\tilde{T}$ between configurations. To determine how and which specific configurations differ from each other, we conducted a *Dunn's test* [31] for each $f \in F$.

Table 4 shows in italics the configurations that, according to the Dunn's test, are *not* significantly different (i.e. are equivalent). We observe that, for *every* $f \in F$, $\text{IDEA}_{\text{adapt}}$ is either the configuration with the minimum $\tilde{T}$ or not significantly different from it. When $f = 360$, $\text{IDEA}_{\text{adapt}}$ has the minimum $\tilde{T}$, and it is statistically different from *all* the other 3 configurations (Figure 9). In this scenario, the deployment of the $\text{IDEA}_{\text{adapt}}$ reduces the $\tilde{T}$ in 13.6%, when compared with not using rescue robots. As seen in Table 4, both $\text{IDEA}_{\text{social}}$ and $\text{IDEA}_{\text{self}}$ are statistically equivalent to the best performer for multiple values of $f \in F$. In contrast, the no-robot support configuration is only equivalent to the best performer in a single instance ($f = 540$), along with *all* the other IDEA configurations.

Overall, our results suggest that IDEA, in its multiple configurations, *does* reduce evacuation time. Our analysis shows that $\text{IDEA}_{\text{adapt}}$ outperforms $\text{IDEA}_{\text{social}}$ and $\text{IDEA}_{\text{self}}$. We posit that this advantage is a consequence of the identity-aware game theoretic models within $\text{IDEA}_{\text{adapt}}$ (section 4). Informed by these models, the $\text{IDEA}_{\text{adapt}}$ robot can decide when it is convenient to request survivor support or contact the first-responder team.

While $\text{IDEA}_{\text{adapt}}$ is *always* statistically indistinguishable from the best performer for every $f \in F$, Table 4 suggest $\text{IDEA}_{\text{adapt}}$ contribution is more impactful for medium-severity disasters, with fall length $180 < f < 420$.

In non-severe disasters ($f < 210$) or disasters with high-complexity ($f > 420$), other IDEA configurations may have more impact than $\text{IDEA}_{\text{adapt}}$. We also observe this pattern in our experiments on environmental and internal factors (subsection 7.2). We posit that in non-severe disasters, the number of victims is low enough that uncoordinated survivors and first-responders can deal with them efficiently, without the mediation of an $\text{IDEA}_{\text{adapt}}$ robot. In contrast, in high-complexity disasters, the number of victims is such that first-responders and survivors are not available to attend $\text{IDEA}_{\text{adapt}}$'s requests for assistance. It is in medium-severity disasters that the deployment of $\text{IDEA}_{\text{adapt}}$ robots make a difference.

7.2 Emergency Severity, Robot Capabilities, and Performance

In this section, we investigate how environmental and internal factors affect the performance of IDEA systems. We postulate that the nature of the emergency and the autonomous system capabilities can affect the contribution of our rescue robot.

We focus on the severity of the emergency as an environmental factor. In the original IMPACT model, when a survivor falls during evacuation, they need a fixed amount of time f to recover before continuing with the evacuation. In IMPACT+ (subsection 6.1), we expose this value as a parameter that we use as a proxy for the severity of the emergency. We assume that, for more severe incidents, the time survivors need to recover is longer.

Regarding internal factors, we explore the impact of the *helping effect for survivors with robot support* (H_{s^r}). Parameter H_{s^r} is another extension we made in IMPACT+. In the original model, when a survivor helps a victim they have no effect in the victim's recovery time. Both survivor and victim *do not* evacuate until the victim is able to continue, according to the value of f. In contrast, in IMPACT+ the IDEA agent intervention reduces the victim's recovery time. Rescue robots can carry fluid delivery mechanisms, oxygen, or triage sensors that survivors can use to assist the victim and accelerate their recovery [91, 99]. In IMPACT+, this effect is represented by the parameter H_{s^r} . The value of H_{s^r} reduces fall duration at every time step, when the IDEA agent mediates survivor assistance to a victim.

Simulation settings. We performed a *sensitivity analysis with factorial design* [76, 103], having fall length (f) and increase in helping effect due to robot support (ΔH_{s^r}) as inputs, and average evacuation time ($\bar{T}$) as output.

Using a factorial design requires to obtain $\bar{T}$ for every element of the Cartesian product $F \times H$, where F is a set of f values and H is a set of values for ΔH_{s^r}. Regarding F, we used the same fall length values as in subsection 7.1. The set of robot-supported helping effect increments is $H = \{2\%, 4\%, \ldots, 38\%, 40\%\}$. We selected $\Delta H_{s^r} < 50\%$ as the upper bound given that it is the helping effect for first-responders (H_f) used in subsection 7.1. Regarding the helping effect for survivors without robot support H_s, we are keeping it constant with a value of $H_s = 2\%$ as in subsection 7.1. For each element of $F \times H$, we obtained T evacuation time samples. A factorial design allows a comprehensive exploration of the parameter space, but it comes at a high computational cost [103]. Due to this, we are only using $|T| = 30$ samples per element of $F \times H$.

Results. Figure 10 shows the average evacuation time $\bar{T}$ for every element of $F \times H$, represented as a heatmap. We observe that, in some instances, lower evacuation times $\bar{T}$ correspond to lower values of fall length f and high values of robot-supported helping effect ΔH_{s^r}. For example, this happens when $f = 240$ and $f = 360$. However, for extreme values of f —either very high or very low– the reduction of $\bar{T}$ associated to a high ΔH_{s^r} loses importance. For example, when $f = 120$ and $f = 510$, there is little difference on $\bar{T}$ when $\Delta H_{s^r} = 2\%$ and $\Delta H_{s^r} = 40\%$.

Given these results, we conclude that the impact of robot capabilities in average evacuation time $\bar{T}$ —represented by the robot-supported helping effect ΔH_{s^r} parameter— is highly dependent on the severity of the disaster, according to fall length f. We posit that, for non-severe disasters with low f, survivors and first-responders are able to cope with the number of victims without the need of advanced capabilities. When disasters have high complexity, and high f values, the advanced capabilities, associated with high values of ΔH_{s^r}, are not enough to evacuate victims quickly. It is in disasters of medium severity where advanced capabilities are more impactful in reducing evacuation time.

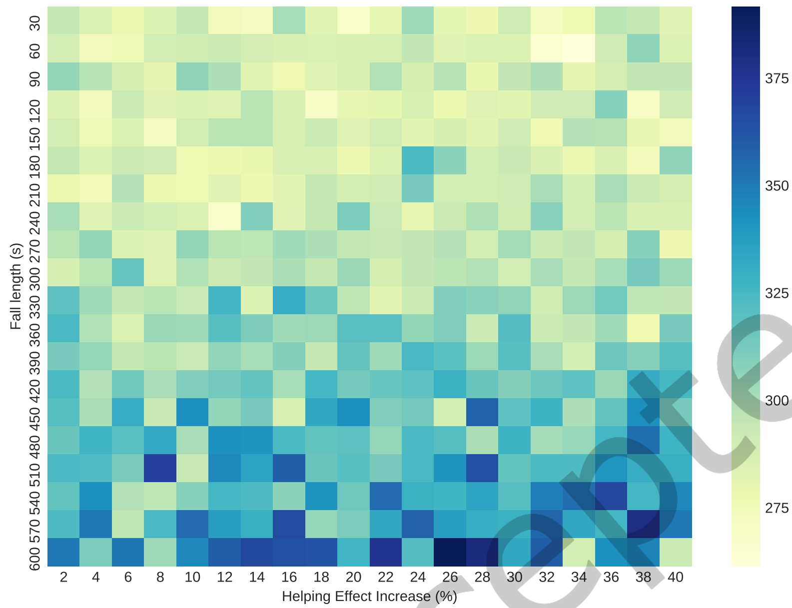

Fig. 10. Sensitivity analysis of input parameters for IDEA systems: 1) fall length f and 2) helping effect increase for survivors with robot support ΔH_{s^r}. In the heatmap, each cell value corresponds to the average evacuation time $\bar{T}$. For mid-range values of f, like $f = 240$ and $f = 360$, $\bar{T}$ decreases when ΔH_{s^r} increases. In contrast, for very low or very high values of f this is not the case.

7.3 Lessons Learned

In this subsection, we describe the key lessons learned from our study:

Modelling human behaviour with group psychology. We provided a technique for modelling human behaviour grounded in social psychology and focussed on group interactions. Modern software systems often involve humans as first-class participants who contribute to decision-making as well as the execution of these decisions [99]. When it comes to modelling and reasoning about humans, many software engineering techniques adopt a behaviourist approach suggesting that every behaviour is a response to a certain stimulus [54], eventually with some uncertainty or probabilities [17, 81]. While these approaches provide a starting point for specifying and reasoning about human behaviour, they are often restrictive in addressing its complexity, as well as the different ways that human groups interact.

System interactions with groups and individuals. Within the human-system cooperation field, IDEA systems cooperate with human groups, using social identity. A significant body of research has been dedicated to designing systems that interact with humans, initially considering interactions with individuals (dyadic interactions) but increasingly considering interactions with groups (non-dyadic interactions) [112]. In software engineering, a lot of the effort was on engineering human-in-the-loop and human-on-the loop systems [17, 25, 81]. Lately, the research community is moving towards human-machine teaming [10, 23], albeit focusing on trained human operators and users. We believe our approach is relevant for engineering human-centred machine learning systems [19] and human-AI collaboration in software engineering [84].

Exemplars of human-system cooperation. Finally, we developed an exemplar for engineering systems that enable cooperation between humans and autonomous systems. Software engineering research not only requires methodological, technical and theoretical results, but also convincing evidence that these results are sound [116]. Exemplars are well-suited for validation, studying relevant problems, and also as a medium for education. Exemplars have been collected and established in various areas of engineering software-intensive systems, e.g., in requirements engineering [34], software and system evolution [134], software product-line engineering [85], and self-adaptive and self-managing systems [1]. To the best of our knowledge, there is no exemplar for engineering identity-aware human-system cooperation. Our exemplar is based on an agent-based model validated empirically by social psychologist in real-world emergencies, and augmented to support the validation of software engineering methods.

7.4 Evaluation Limitations

We based our evaluation on an independently developed simulation model, which was engineered by domain experts in social psychology and emergency evacuations [131]. However, using simulation models poses an internal validity threat, given that these models rely on assumptions and simplifications made by the simulation designer [24]. For example, in IMPACT+ there is an underlying assumption that survivors and first-responders only reduce the victim's recovery time. However, there is a non-negligible probability —especially in emergencies— that complications or mishandling deteriorate the victim's condition.

Given IDEA's focus on identity-driven adaptation, the simulation model emphasises representing social agents and their interactions, while neglecting other aspects that affect system performance. For instance, the capabilities of a rescue robot can be affected by sensor failure, environmental noise, or reduced visibility due to smoke. We built IMPACT+ using the NetLogo agent-based platform. Agent-based models are well-suited for representing social interactions [150], and they are widely used for evaluating emergency evacuations [21, 53]. However, they struggle to model complex environmental factors. Other simulation platforms, like Gazebo, integrate with physics engines that are better suited for environmental modelling [69]. However, the engineering effort of developing such models for social agents is significant. We plan to test IDEA in a high-fidelity simulation environment in future work, with the goal of a later deployment in a physical robot.

Regarding construct validity, using evacuation time as the only metric of robot performance can constitute a threat, given that it excludes, for example, number of casualties. We adopted evacuation time as a metric, given its popularity within the emergency evacuation literature [83] and its availability in the IMPACT model [131]. Capturing additional metrics require further modifications to the agent-based model, that we plan to address in future extensions of this work.

In addition, we will explore further measures of performance beyond evacuation time such as 1) *empowerment,* defined as the ability to enhance the agency and capabilities of zero responders to respond to emergencies, 2) *connectedness,* defined as the ability to encourage interactions between different stakeholders and autonomous systems, and 3) *sociotechnical resilience,* defined as the ability to maintain an effective, responsible, and inclusive emergency response under unanticipated events and disruption. In future work, we plan to conduct lab-based experiments to evaluate IDEA system's impact in the evacuation outcome, in a controlled environment. We plan to measure evacuation performance based on evacuation time and number of successful evacuations [6]. We will also monitor connectedness via the identity fusion pictorial scale [123], which captures the degree of integration between the self and the group. We will measure empowerment on the psychological empowerment scale [121] that taps into cognitive, emotional, and behavioural dimensions of community participation.

8 DEPLOYING IDEA SYSTEMS

In this section, we describe the considerations and challenges for bringing IDEA into real-world systems.

Hardware and Software. IDEA is our proposed architecture, that can be realised in concrete self-adaptive systems in multiple ways. At software level, each IDEA component can be decomposed in fine-level components with specific responsibilities. For instance, in a ROS implementation of Figure 5's scenario, the Game Selector can contain a node in charge of transforming audio data into text, and a different one for estimating survivor distance from LIDAR and camera inputs. At the hardware level, some IDEA components can be deployed on-board into the system's hardware, while others can be delegated to external environments. For example, the Game Solver and the Identity Estimator can be deployed in the cloud as web services and accessed remotely via HTTP. Or the sensor data processed by the Game Selector can be enriched with external sources, like video streams from camera drones. Engineering teams developing IDEA systems need to tailor their implementation to their context. Memory, processing power, and network access vary greatly between an on-board deployment on an NVIDIA Jetson Nano or on an external ground-control station.

The control loop architecture within IDEA enable systems that operate among humans to adapt to their identity-driven behaviours. IDEA components, described in section 5, work together to suggest a system action given the perceived social identity of a human in its proximity. In real-world deployments, systems have other adaptation concerns besides identity-driven behaviours, that can be handled by one or many feedback control loops [41, 140]. For example, commercial drones nowadays incorporate mechanisms for responding to low battery, signal loss, or geo-fence breaches. And other controllers in the literature can adapt to software errors and hacking attacks. [99]. Hence, at deployment, IDEA components should constitute a control loop within a wider autonomic manager, interacting with control loops that handle other adaptation concerns.

The game-theoretic models that inform IDEA's decision making require special care. Engineering teams need to instantiate our modelling framework, described in subsection 4.2, into a concrete game tree like the one in Figure 3, that explains human-human and human-system interactions. Many research disciplines, like behavioural game-theory, actively study the alignment of game-theoretic models to human behaviour [147]. Successful real-world applications of game-theory often need laboratory experiments and have whole teams of game-theorists behind them [146]. While it is not in the scope of this paper to provide a comprehensive primer on game-theoretic modelling, we would recommend engineering teams to: 1) keep models simple to maximise the likelihood of rational choice [146], 2) explore the impact of social norms and incentives in payoff function design [133, 146], and 3) iteratively develop and validate model candidates, relying on real-world experiments, data analysis, and simulation [43, 47].

Challenges. The identity estimator (subsection 5.2), arguably the most important element of IDEA systems, is a machine learning component. Incorporating machine learning systems within self-adaptive systems raises a scalability challenge [49]. In the search-and-rescue example, we need the identity estimator to transform sensor readings into identity predictions within milliseconds. Otherwise, the human interacting with the robot might lose interest and leave. Also, training the identity estimator can take significant amount of time and compute, considering identity markers can be sourced from large quantities of audio and video. In this work, we could not explore this scalability considerations, given that our evaluation was done entirely on a mobile workstation, using an agent-based model. There are approaches within the practitioner community that deal with scaling machine learning systems [15], that we plan to explore in future work.

Another scalability challenge is related to the size of identity-driven game theoretic models. The model size used in section 4 is modest: two agents, two actions per player, and two candidate identities for the human agent. Each of these game parameters have a big impact on model size. Adopting game abstraction techniques can potentially reduce model size, while keeping its prediction properties [109]. We can also investigate the effect of increasing the autonomous system's computing resources, or delegating some computations to an external server.

Finally, in this paper, we focused on IDEA systems that reason about the interaction between one rescue robot, one survivor, and one first responder. While IDEA can be used in situations with multiple robots and humans, it does not yet support non-dyadic interventions. There exists a large body of work focusing on non-dyadic human-system interactions [111] (m systems, n humans, $m, n > 1$) , albeit focusing on design rather than reasoning. In future work, we plan to extend IDEA by reasoning about groups of humans and their cooperation with multiple robotic systems. By focusing on social identity, IDEA paves the way to reasoning about groups and group membership. In addition, building on game theory, we can exploit multi-agent reinforcement learning techniques to build adaptation strategies in uncertain environments [75]. This is particularly relevant in search-and-rescue scenarios, where multi-robot interventions are increasingly used and require coordination with diverse human operators [100].

9 LIMITATIONS AND FUTURE WORK

This section discusses the assumptions we made in order to implement and empirically evaluate our proposed software architecture. It also discusses the limitations of the approach and recommends next steps.

Modelling human behaviour. IDEA systems reason about human behaviour using identity-driven game-theoretic models (subsection 4.2). As with any modelling framework, game-theoretic predictions only hold if their assumptions are met. In games of incomplete information —like the one in Figure 3— these assumptions include: 1) players are rational, and try to maximise their payoff, 2) players know their own type, but not the one of their opponents, and 3) the probability distribution over types is common knowledge [79]. While these assumptions enable making behavioural predictions, they may not hold in real-world situations. In future work, we plan to explore game-theoretic frameworks that relax some of these assumptions. For example, approaches like bounded rationality relax the rationality assumptions by imposing limits on a player's attention, information, and mental capabilities [57].

Besides the assumptions within game theory, we made further simplifications in IDEA models to keep their size tractable. For example, the model in Figure 3 assumes survivor reaction is only a function of shared identity adoption, with only two possible reactions to the rescue robot proposal. This model does not support survivors having other behavioural drivers like mental state, or adopting actions like vandalising the rescue robots. Simplifying assumptions are a common and necessary modelling technique, but there is always a risk of oversimplifying to the extent of making models irrelevant ("all models are wrong, but some are useful" [14]). While our simulation-based evaluation suggests this is not the case, in future work we want to explore if IDEA systems also perform well in higher-fidelity simulation contexts.

In previous work, we relied on using language to infer social identity in emergencies [70]. We recognise that there are multiple and diverse markers for social identity. For example, they can be associated with stable social categories, such as gender, age or cultural background [38], such as within the IMPACT model. Markers can also be part of a person's discourse, their clothing, or even their accent. For example, a football-supporter identity can be expressed via jerseys, scarfs, or face paint. Identity markers are also related to how close people are in space, they way people walk, or how they speak [70]. Similarly, disability has been increasingly construed as an identity [36], making it an important factor during evacuations. We plan to improve our identity estimator by considering multiple markers and adaptively composing their outputs according to the context. While identity inference is uncertain, we aim to improve it to approach human accuracy, leveraging the data and analysis carried out during emergencies [94].

Application Domains. Emergencies are heterogeneous events. Among others, they can differ on number of casualties, geographical area affected, duration, and cause [53]. As described in section 8, we explored the parameter space covering emergency scenarios of diverse severity levels, supported by robots with different capabilities. However, due to compute and time limitations, there is a portion of the parameter space that we did not cover in this work. Our simulation code exposes severity levels and robot capabilities as user-provided values, so in future work we can increase our coverage of the parameter space.

Given the domain of IMPACT+, there is an external validity threat about the generalisation of our findings to other identity-related situations. This study is clearly focused on identity-driven behaviours *within* emergency evacuations, and the application of IDEA to other contexts requires further study. Social identity has been applied in multiple domains, including security [102], privacy [16], and healthcare [122]. For example, it has been shown that during a health emergency, bystanders exhibiting helping behaviours can increase the survival rate by providing cardiopulmonary resuscitation (CPR). In 2142 emergency medical situations, 22.9% of the victims survived when bystander CPR was administered, whereas only 14.6% survived when bystander CPR was not administered [106]. More recently, it has been found that bystanders are willing to use automatic external defibrillators to help victims. When delivered by a drone, which reduces arrival time, the expected survival rate due to bystander-drone collaboration has been doubled in the emergencies tested [12]. While in this paper we focus on emergency evacuations, in future work we will explore the application of IDEA systems to other domains. For example, IDEA systems can be used in healthcare to support adaptations that consider the patient and their support groups [9]; or can facilitate software-mediated cooperation between robotics engineers [43].

10 RELATED WORK

IDEA's goal is to maximise the effectiveness of human-system collaboration, by equipping autonomous systems with an understanding of social identity. Other work enables human-system collaboration by modifying the robot's appearance [126], or by tailoring the language that defines the human-system relationship [73]. In that sense, IDEA is analogous to the framework developed by Hoffman and Breazeal, where robots communicate with humans, split a task and negotiate which actions to take [55]. In comparison, IDEA's models are limited to representing a single request for help and the human's response to it. It is possible to extend IDEA's modelling capabilities to support complex interactions, including multiple tasks and negotiation rounds. We leave the exploration of this idea to future work.

We start this section discussing work related to three key characteristics of the IDEA software architecture: 1) using psychology for autonomous system design, 2) autonomous system's models of human preferences, and 3) using sensor data to predict human disposition for cooperation. IDEA is an *adaptive* architecture for autonomous systems, so we finish with a review of self-adaptive systems in general.

Psychology and autonomous system design. We use social identity theory to provide autonomous systems with an understanding of group behaviour. Like us, other researchers applied social identity theory within computing. For example, Seering *et al.* proposed a set of guidelines for incorporating social identity techniques into computer-supported cooperative work systems [115]. Calikli *et al.* used models of social identity to extract privacy norms from online social networks [16]. And Rauf *et al.* explained developer's attitude towards security concerns using social identity [102]. This body of work focuses on understanding and facilitating interactions between group members. In this paper, our goal is to enable cooperation between group members and autonomous systems.

Other researchers in computing rely on different psychological theories to understand group behaviour. For example, Ahrndt *et al.* applied the Five-Factor Model to learn personalities within a team [4]. Schwarting *et al.* used Social Value Orientation to quantify how much individuals value personal rewards in relation to group-level rewards [113]. Smaldino *et al.* used the Optimal Distinctiveness Theory to model group formation [119]. Vinitsky *et al.* developed an architecture for cooperating agents emulating human social norms [133]. And

Pietraszewski developed a new computational theory of social groups, suitable for its adoption by artificial intelligence systems [95]. Our focus on social identity was driven by our application to rescue robotics. There is an extensive body of research on the role of social identity in emergency evacuations [135]. This includes simulation models suitable for validating our approach [37, 38, 131].

Modelling human preferences. In IDEA, we incorporate social identity insights into game-theoretic models, to inform autonomous systems about the outcomes preferred by humans (section 4). Other researchers use different modelling techniques for this purpose. For instance, Gal *et al.* model human preferences using utility functions, that can be learned after repeated human-agent interactions [40]. Besides its support for uncertain preferences, our adoption of game theory was influenced by its track record on modelling human behaviour during evacuations [50, 62, 63]. Game theory has also been used extensively to model interactions between humans and autonomous systems [74, 82, 136, 148]. Namely, Messeri *et al.* used game theory to obtain adaptation strategies that minimise stress, while maximising the productivity of a human-robot team [88]. Also, Schwarting *et al.* proposed autonomous vehicles that use game-theoretic models for predicting the behaviour of human drivers [113].

Predicting human-robot cooperation. In this work, we rely on identity markers —informed by social psychology research— to infer the probability of cooperation, given a shared identity. Other researchers focus on learning features correlated with cooperation, from people's interaction data. For example, Kantharaju and Pelachaud analysed a patient consultation corpus, extracting non-verbal cues —like facial expressions— from cohesive groups [66]. Schmuck and Celiktutan inferred group membership from position and orientation data extracted from images [110]. Hung and Chittaranjan proposed predicting roles within a group using features extracted from audio samples [61]. Cherubini *et al.* developed a robotic helper for human operators inserting screws, that predicts the operator intentions using data from a Kinect sensor [20]. Huang and Mutlu also developed an intention prediction mechanism, based on glance data collected from eye-tracking glasses [59].

This body of work is complementary to our approach: if facial expressions, audio features, or orientation are indicative of group cooperation, we can incorporate them within the Identity Estimator component (subsection 5.1). In future work, we will explore additional sensors —like eye-tracking ones— and their potential to infer *groupiness*. Also, we will continue our exploration of identity markers and their impact on robot performance.

Self-Adaptive Systems. The need for self-management and self-adaptation is inherent to autonomous systems: they are long-lived, continuously running systems that interact with the environment and humans in ways that cannot be fully anticipated at design time, and continuously evolve at runtime. There are a plethora of architectures and techniques for engineering self-adaptive systems: we refer the interested reader to a survey of those techniques [138]. However, a long-lasting challenge for engineering self-adaptation is dealing with uncertainty, which can span multiple areas, including requirements, environments, other systems, and humans [139]. In this paper, we are interested in the uncertainty of human behaviour.

Numerous self-adaptive systems are complex sociotechnical systems. Humans interacting with them are not merely providers of system input and consumers of artefacts produced by them. They are first-class participants, and the system relies on human contributions to make decisions and executing them [99]. This requires self-adaptation methods to consider human participants during their lifecycle: from the monitoring and analysis of the system and its environment, to the synthesis of adaptation plans and the execution of these plans. Building on the work of Schneiders [112], we can group autonomous systems along two dimensions: the number of autonomous agents (single or multiple), and whether humans interact with them. Accordingly, systems can take one of four forms: 1) a single autonomous agent (e.g., UAV); 2) a group of autonomous agents (e.g., swarms); 3) an autonomous agent with a human (e.g., AI in healthcare); and 4) a group of autonomous agents with humans (e.g., emergency situations). In this paper, we focus on the last case, and we refer the interested reader to Abeywickrama *et al.* paper for a description of the other areas and related challenges [2].

Different theories of human behaviour explain human actions in different ways, by focussing on different determinants of human behaviour. For example, a behaviourist approach suggests that every behaviour is a response to a certain stimulus [54]. Under this approach, to deal with uncertainty we can use probabilistic behavioural modelling [17, 81]. This approach is widely adopted for engineering self-adaptive systems with *humans-in-the-loop*, i.e., where humans make decisions at key points [25]. Research has also emphasised the need for social adaptation, where the autonomous system analyses users' feedback and updates its behaviour to best satisfy the requirements in the given context [7]. For example, when applied to emergency management, the autonomous system can model human behaviour and decision-making probabilistically and plan accordingly [65]. With increasing autonomy, a *human-on-the-loop* approach considers humans as "strategic" supervisors, while the autonomous systems are responsible for planning specific tasks autonomously [35, 80]. For example, when applied to emergency management, a first responder may monitor the plans and actions of the autonomous systems and intervene when necessary.

While human-in-the loop and human-on-the-loop approaches provide a starting point for specifying and reasoning about human behaviour, they are often restrictive in addressing the different ways that human behaviour develops during their interaction with software. To grasp that humans have purposes and goals that affect each other, the concept of joint-action can be introduced as "a social interaction whereby two or more individuals coordinate their actions in space and time to bring about change in the environment" [114]. This approach suggests an interplay between humans and self-adaptive systems, such that what matters is not only how the self-adaptive systems understands itself but also how humans understand the way the system behaves [51]. In this context, the work on *human-machine teaming* [23] is the closest to ours. In particular, the proposed $MAPE\text{-}K_{HMT}$ framework uses runtime models to augment the monitoring, analysis, planning, and execution phases to support humans-machines teaming, emphasising the bidirectional relationship between humans and autonomous systems. When applied in emergency situations, drones and first responders continuously interact with one another and work together, exchanging information to best respond to the emergency. However, in human-machine teaming the human collaborators are trained, belong to the same group, and share the same goals as the drones. In our paper, we seek to ensure cooperation when the humans belong to different groups and may have different goals. Figure 11 summarises those different perspectives, mentions some approaches, and illustrates what it means in the case of emergency scenarios. The figure shows from bottom to top an increase in human interaction and cooperation.

11 CONCLUSION

Designing autonomous systems to cooperate with humans is not simply an engineering problem, but also requires reasoning about human behaviour and group dynamics based on psychological research. In this paper, we build upon social identity theory to enable cooperation between autonomous systems and humans. We demonstrated the approach by developing an identity-aware rescue robot, which leverages existing research on social identity in emergency situations. In particular, we believe that this paper contributes to software engineering research in the following topics:

We presented a novel approach for designing software-intensive adaptive systems, capable of cooperating with humans based on their group identity. To bring our approach to other domains, engineering teams and social psychologists need to work together to determine the social identities relevant to human-robot cooperation in different contexts. This investigation must include discovering identity markers, and how these markers will be processed by the autonomous system. We believe IDEA is a step towards more resilient and cooperative

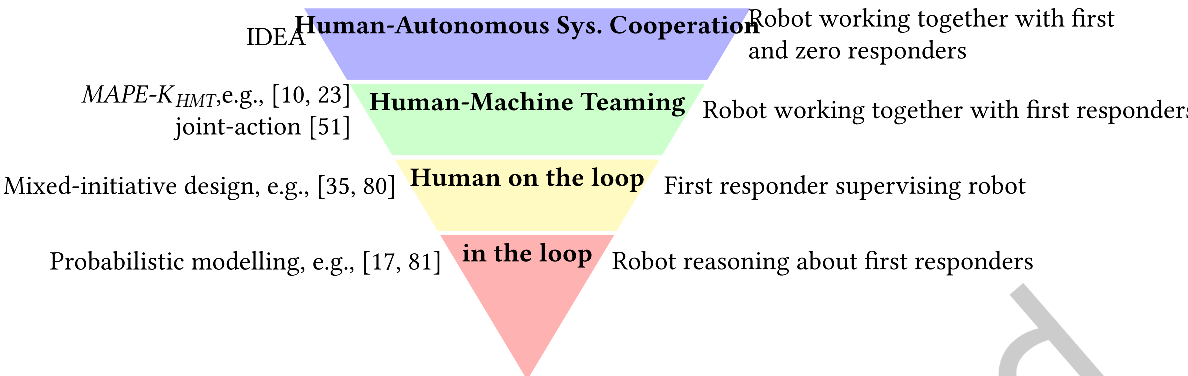

Fig. 11. Overview of self-adaptive approaches to human cooperation

autonomous systems, and we welcome participation and inquiries from users, researchers, and other stakeholders within and beyond the boundaries of the emergency domain.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant numbers EP/V026747/1, EP/R013144/1]; and Science Foundation Ireland [grant number 13/RC/2094_P2].

REFERENCES

[1] 2019. Software Engineering for Self-Adaptive Systems Exemplars. https://www.hpi.uni-potsdam.de/giese/public/selfadapt/exemplars/.

[2] Dhaminda B. Abeywickrama, Amel Bennaceur, Greg Chance, Yiannis Demiris, Anastasia Kordoni, Mark Levine, Luke Moffat, Luc Moreau, Mohammad Reza Mousavi, Bashar Nuseibeh, Subramanian Ramamoorthy, Jan Oliver Ringert, James Wilson, Shane Windsor, and Kerstin Eder. 2024. On Specifying for Trustworthiness. *Commun. ACM* 67, 1 (2024), 98–109. https://doi.org/10.1145/3624699

[3] Håkon B Abrahamsen. 2015. A remotely piloted aircraft system in major incident management: concept and pilot, feasibility study. *BMC emergency medicine* 15 (2015), 1–12.

[4] Sebastian Ahrndt, Benjamin Breitung, Johannes Fähndrich, and Sahin Albayrak. 2015. Predictability in human-agent cooperation: adapting to humans' personalities. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing, Salamanca, Spain, April 13-17, 2015*, Roger L. Wainwright, Juan Manuel Corchado, Alessio Bechini, and Jiman Hong (Eds.). ACM, 474–479. https://doi.org/10.1145/2695664.2695702

[5] George A. Akerlof and Rachel E. Kranton. 2000. Economics and Identity*. *The Quarterly Journal of Economics* 115, 3 (08 2000), 715–753. https://doi.org/10.1162/003355300554881 arXiv:https://academic.oup.com/qje/article-pdf/115/3/715/5174909/115-3-715.pdf

[6] Yasser M Alginahi, Mohammed Mudassar, Muhammad Nomani Kabir, and Omar Tayan. 2019. Analyzing the crowd evacuation pattern of a large densely populated building. *Arabian Journal for Science and Engineering* 44 (2019), 3289–3304.

[7] Malik Almaliki, Funmilade Faniyi, Rami Bahsoon, Keith Phalp, and Raian Ali. 2014. Requirements-Driven Social Adaptation: Expert Survey. In *Requirements Engineering: Foundation for Software Quality - 20th International Working Conference, REFSQ 2014, Essen, Germany, April 7-10, 2014. Proceedings (Lecture Notes in Computer Science, Vol. 8396)*, Camille Salinesi and Inge van de Weerd (Eds.). Springer, 72–87. https://doi.org/10.1007/978-3-319-05843-6_6

[8] Bart Baesens. 2014. *Analytics in a big data world: The essential guide to data science and its applications.* John Wiley & Sons.

[9] Amel Bennaceur, Avelie Stuart, Blaine A. Price, Arosha K. Bandara, Mark Levine, Linda Clare, Jessica Cohen, Ciaran McCormick, Vikram Mehta, Mohamed Bennasar, Daniel Gooch, Carlos Gavidia-Calderon, Anastasia Kordoni, and Bashar Nuseibeh. 2023. Socio-Technical Resilience for Community Healthcare. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems, TAS 2023, Edinburgh, United Kingdom, July 11-12, 2023.* ACM, 26:1–26:6. https://doi.org/10.1145/3597512.3599720

[10] Marcello M. Bersani, Matteo Camilli, Livia Lestingi, Raffaela Mirandola, Matteo G. Rossi, and Patrizia Scandurra. 2023. Towards Better Trust in Human-Machine Teaming through Explainable Dependability. In *20th International Conference on Software Architecture, ICSA 2023 - Companion, L'Aquila, Italy, March 13-17, 2023.* IEEE, 86–90. https://doi.org/10.1109/ICSA-C57050.2023.00029

[11] Andreas Birk and Stefano Carpin. 2006. Rescue robotics - A crucial milestone on the road to autonomous systems. *Adv. Robotics* 20, 5 (2006), 595–605. https://doi.org/10.1163/156855306776985577
[12] Brittany Bogle, Wayne D Rosamond, Kyle T Snyder, and Jessica K Zègre-Hemsey. 2019. The case for drone-assisted emergency response to cardiac arrest: an optimized statewide deployment approach. *North Carolina medical journal* 80, 4 (2019), 204.
[13] Tibor Bosse, Mark Hoogendoorn, Michel C. A. Klein, Jan Treur, C. Natalie van der Wal, and Arlette van Wissen. 2013. Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions. *Auton. Agents Multi Agent Syst.* 27, 1 (2013), 52–84. https://doi.org/10.1007/s10458-012-9201-1
[14] GE Box. 1979. All models are wrong, but some are useful. *Robustness in Statistics* 202, 1979 (1979), 549.
[15] Henrik Brink, Joseph Richards, and Mark Fetherolf. 2016. *Real-World Machine Learning* (1st ed.). Manning Publications Co., USA.
[16] Gül Çalikli, Mark Law, Arosha K. Bandara, Alessandra Russo, Luke Dickens, Blaine A. Price, Avelie Stuart, Mark Levine, and Bashar Nuseibeh. 2016. Privacy dynamics: learning privacy norms for social software. In *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS@ICSE 2016, Austin, Texas, USA, May 14-22, 2016*. ACM, 47–56. https://doi.org/10.1145/2897053.2897063
[17] Javier Cámara, Gabriel Moreno, and David Garlan. 2015. Reasoning about human participation in self-adaptive systems. In *2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*. IEEE, 146–156.
[18] Holly Carter, Dale Weston, Richard Amlôt, SA Haslam, and T Cruwys. 2020. Managing crowds in crises. *Together apart: The psychology of COVID-19* 13 (2020), 88–92.
[19] Stevie Chancellor. 2023. Toward Practices for Human-Centered Machine Learning. *Commun. ACM* 66, 3 (2023), 78–85. https://doi.org/10.1145/3530987
[20] Andrea Cherubini, Robin Passama, Arnaud Meline, André Crosnier, and Philippe Fraisse. 2013. Multimodal control for human-robot cooperation. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, November 3-7, 2013*. IEEE, 2202–2207. https://doi.org/10.1109/IROS.2013.6696664
[21] Keith Christensen and Yuya Sasaki. 2008. Agent-Based Emergency Evacuation Simulation with Individuals with Disabilities in the Population. *J. Artif. Soc. Soc. Simul.* 11, 3 (2008). http://jasss.soc.surrey.ac.uk/11/3/9.html
[22] Civil Contingencies Secretariat. 2004. Civil Contingencies Act 2004. https://www.ifrc.org/docs/idrl/782EN.pdf
[23] Jane Cleland-Huang, Theodore Chambers, Sebastian Zudaire, Muhammed Tawfiq Chowdhury, Ankit Agrawal, and Michael Vierhauser. 2023. Human-Machine Teaming with Small Unmanned Aerial Systems in a MAPE-K Environment. *ACM Trans. Auton. Adapt. Syst.* (sep 2023). https://doi.org/10.1145/3618001 Just Accepted.
[24] Breno Bernard Nicolau de França and Guilherme Horta Travassos. 2016. Experimentation with dynamic simulation models in software engineering: planning and reporting guidelines. *Empir. Softw. Eng.* 21, 3 (2016), 1302–1345. https://doi.org/10.1007/s10664-015-9386-4
[25] Rogério de Lemos. 2020. Human in the loop: what is the point of no return?. In *SEAMS '20: IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, Seoul, Republic of Korea, 29 June - 3 July, 2020*, Shinichi Honiden, Elisabetta Di Nitto, and Radu Calinescu (Eds.). ACM, 165–166. https://doi.org/10.1145/3387939.3391597
[26] Kay Deaux. 1993. Reconstructing social identity. *Personality and social psychology bulletin* 19, 1 (1993), 4–12.
[27] John Drury. 2018. The role of social identity processes in mass emergency behaviour: An integrative review. *European Review of Social Psychology* 29, 1 (2018), 38–81. https://doi.org/10.1080/10463283.2018.1471948 arXiv:https://doi.org/10.1080/10463283.2018.1471948
[28] John Drury, Holly Carter, Chris Cocking, Evangelos Ntontis, Selin Tekin Guven, and Richard Amlôt. 2019. Facilitating collective psychosocial resilience in the public in emergencies: Twelve recommendations based on the social identity approach. *Frontiers in public health* 7 (2019), 141.
[29] John Drury, Chris Cocking, and Steve Reicher. 2009. Everyone for themselves? A comparative study of crowd solidarity among emergency survivors. *British Journal of Social Psychology* 48, 3 (2009), 487–506. https://doi.org/10.1348/014466608X357893 arXiv:https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/014466608X357893
[30] John Drury, Christopher Cocking, and Stephen D. Reicher. 2009. The Nature of Collective Resilience: Survivor Reactions to the 2005 London Bombings. *International journal of mass emergencies and disasters* 27 (2009), 66–95.
[31] Olive Jean Dunn. 1964. Multiple Comparisons Using Rank Sums. *Technometrics* 6, 3 (1964), 241–252. https://doi.org/10.1080/00401706.1964.10490181 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/00401706.1964.10490181
[32] Alice H. Eagly and Maureen Crowley. 1986. Gender and Helping Behavior. A Meta-Analytic Review of the Social Psychological Literature. *Psychological Bulletin* 100, 3 (Nov. 1986), 283–308. https://doi.org/10.1037/0033-2909.100.3.283
[33] Milan Erdelj, Enrico Natalizio, Kaushik R Chowdhury, and Ian F Akyildiz. 2017. Help from the sky: Leveraging UAVs for disaster management. *IEEE Pervasive Computing* 16, 1 (2017), 24–32.
[34] Martin S Feather, Stephen Fickas, Anthony Finkelstein, and Axel Van Lamsweerde. 1997. Requirements and specification exemplars. *Automated Software Engineering* 4, 4 (1997), 419–438.
[35] Joel E. Fischer, Chris Greenhalgh, Wenchao Jiang, Sarvapali D. Ramchurn, Feng Wu, and Tom Rodden. 2021. In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent. *Concurr. Comput. Pract. Exp.* 33, 8 (2021). https://doi.org/10.1002/cpe.4082

[36] Anjali J Forber-Pratt, Dominique A Lyew, Carlyn Mueller, and Leah B Samples. 2017. Disability identity development: A systematic review of the literature. *Rehabilitation psychology* 62, 2 (2017), 198.

[37] Daniel Formolo, Tibor Bosse, and C. Natalie van der Wal. 2018. Studying the Impact of Trained Staff on Evacuation Scenarios by Agent-Based Simulation. In *Social Informatics - 10th International Conference, SocInfo 2018, St. Petersburg, Russia, September 25-28, 2018, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 11186)*, Steffen Staab, Olessia Koltsova, and Dmitry I. Ignatov (Eds.). Springer, 85–96. https://doi.org/10.1007/978-3-030-01159-8_8

[38] Daniel Formolo and C. Natalie van der Wal. 2017. Simulating Collective Evacuations with Social Elements. In *Computational Collective Intelligence - 9th International Conference, ICCCI 2017, Nicosia, Cyprus, September 27-29, 2017, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 10448)*, Ngoc Thanh Nguyen, George A. Papadopoulos, Piotr Jedrzejowicz, Bogdan Trawinski, and Gottfried Vossen (Eds.). Springer, 160–171. https://doi.org/10.1007/978-3-319-67074-4_16

[39] Bruno S. Frey, David A. Savage, and Benno Torgler. 2010. Interaction of natural survival instincts and internalized social norms exploring the Titanic and Lusitania disasters. *Proceedings of the National Academy of Sciences* 107, 11 (2010), 4862–4865. https://doi.org/10.1073/pnas.0911303107 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.0911303107

[40] Ya'akov Gal, Avi Pfeffer, Francesca Marzo, and Barbara J. Grosz. 2004. Learning Social Preferences in Games. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA*, Deborah L. McGuinness and George Ferguson (Eds.). AAAI Press / The MIT Press, 226–231. http://www.aaai.org/Library/AAAI/2004/aaai04-036.php

[41] David Garlan, Shang-Wen Cheng, An-Cheng Huang, Bradley R. Schmerl, and Peter Steenkiste. 2004. Rainbow: Architecture-Based Self-Adaptation with Reusable Infrastructure. *Computer* 37, 10 (2004), 46–54. https://doi.org/10.1109/MC.2004.175

[42] Carlos Gavidia-Calderon, Amel Bennaceur, Anastasia Kordoni, Mark Levine, and Bashar Nuseibeh. 2022. What Do You Want From Me? Adapting Systems to the Uncertainty of Human Preferences. In *2022 IEEE/ACM 44th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*. 126–130. https://doi.org/10.1109/ICSE-NIER55298.2022.9793539

[43] Carlos Gavidia-Calderon, Amel Bennaceur, Tamara Lopez, Anastasia Kordoni, Mark Levine, and Bashar Nuseibeh. 2023. Meet your Maker: A Social Identity Analysis of Robotics Software Engineering. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems, TAS 2023, Edinburgh, United Kingdom, July 11-12, 2023*. ACM, 44:1–44:5. https://doi.org/10.1145/3597512.3600206

[44] Carlos Gavidia-Calderon, Anastasia Kordoni, Amel Bennaceur, Mark Levine, and Bashar Nuseibeh. 2022. *Identity-Aware Autonomous Systems : Agent-Based Model*. https://github.com/cptanalatriste/robot-assisted-evacuation

[45] Carlos Gavidia-Calderon, Anastasia Kordoni, Amel Bennaceur, Mark Levine, and Bashar Nuseibeh. 2022. *Identity-Aware Autonomous Systems : Code Repository*. https://github.com/cptanalatriste/wdywfm-adaptive-robot

[46] Carlos Gavidia-Calderon, Anastasia Kordoni, Amel Bennaceur, Mark Levine, and Bashar Nuseibeh. 2023. *Code Repository: Are we in this together? Embedding social identity detection in drones improves emergency coordination*. https://github.com/cptanalatriste/transformer-type-estimator

[47] Carlos Gavidia-Calderon, Federica Sarro, Mark Harman, and Earl T. Barr. 2020. Game-theoretic analysis of development practices: Challenges and opportunities. *J. Syst. Softw.* 159 (2020). https://doi.org/10.1016/j.jss.2019.110424

[48] Carlos Gavidia-Calderon, Federica Sarro, Mark Harman, and Earl T. Barr. 2021. The Assessor's Dilemma: Improving Bug Repair via Empirical Game Theory. *IEEE Trans. Software Eng.* 47, 10 (2021), 2143–2161. https://doi.org/10.1109/TSE.2019.2944608

[49] Omid Gheibi, Danny Weyns, and Federico Quin. 2020. Applying Machine Learning in Self-adaptive Systems: A Systematic Literature Review. *ACM Trans. Auton. Adapt. Syst.* 15, 3 (2020), 9:1–9:37. https://doi.org/10.1145/3469440

[50] Julio E. Godoy, Ioannis Karamouzas, Stephen J. Guy, and Maria L. Gini. 2015. Adaptive Learning for Multi-Agent Navigation. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, Gerhard Weiss, Pinar Yolum, Rafael H. Bordini, and Edith Elkind (Eds.). ACM, 1577–1585. http://dl.acm.org/citation.cfm?id=2773353

[51] Elena Corina Grigore, Kerstin Eder, Anthony G. Pipe, Chris Melhuish, and Ute Leonards. 2013. Joint action understanding improves robot-to-human object handover. In *Proc. of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 4622–4629. https://doi.org/10.1109/IROS.2013.6697021

[52] Cynthia Hardy, Thomas B. Lawrence, and David Grant. 2005. Discourse and Collaboration: The Role of Conversations and Collective Identity. *Academy of Management Review* 30, 1 (2005), 58–77. https://doi.org/10.5465/amr.2005.15281426 arXiv:https://doi.org/10.5465/amr.2005.15281426

[53] Glenn I. Hawe, Graham Coates, Duncan T. Wilson, and Roger S. Crouch. 2012. Agent-based simulation for large-scale emergency response: A survey of usage and implementation. *ACM Comput. Surv.* 45, 1 (2012), 8:1–8:51. https://doi.org/10.1145/2379776.2379784

[54] Joe E Heimlich and Nicole M Ardoin. 2008. Understanding behavior to understand behavior change: A literature review. *Environmental education research* 14, 3 (2008), 215–237.

[55] Guy Hoffman and Cynthia Breazeal. 2004. Collaboration in human-robot teams. In *AIAA 1st intelligent systems technical conference*. 6434.

[56] Eric Holgate, Isabel Cachola, Daniel Preotiuc-Pietro, and Junyi Jessy Li. 2018. Why Swear? Analyzing and Inferring the Intentions of Vulgar Expressions. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium,*

October 31 - November 4, 2018, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 4405–4414. https://aclanthology.org/D18-1471/

[57] Keith J Holyoak and Robert G Morrison. 2012. *The Oxford handbook of thinking and reasoning.* Oxford University Press.

[58] Nick Hopkins and Stephen Reicher. 2021. Mass Gatherings, health, and Well-Being: from risk mitigation to health promotion. *Social Issues and Policy Review* 15, 1 (2021), 114–145.

[59] Chien-Ming Huang and Bilge Mutlu. 2016. Anticipatory Robot Control for Efficient Human-Robot Collaboration. In *The Eleventh ACM/IEEE International Conference on Human Robot Interation, HRI 2016, Christchurch, New Zealand, March 7-10, 2016*, Christoph Bartneck, Yukie Nagai, Ana Paiva, and Selma Sabanovic (Eds.). IEEE/ACM, 83–90. https://doi.org/10.1109/HRI.2016.7451737

[60] Matthew Huggins, Sharifa Alghowinem, Sooyeon Jeong, Pedro Colon-Hernandez, Cynthia Breazeal, and Hae Won Park. 2021. Practical Guidelines for Intent Recognition: BERT with Minimal Training Data Evaluated in Real-World HRI Application. In *HRI '21: ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, March 8-11, 2021*, Cindy L. Bethel, Ana Paiva, Elizabeth Broadbent, David Feil-Seifer, and Daniel Szafir (Eds.). ACM, 341–350. https://doi.org/10.1145/3434073.3444671

[61] Hayley Hung and Gokul Chittaranjan. 2010. The idiap wolf corpus: exploring group behaviour in a competitive role-playing game. In *Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010*, Alberto Del Bimbo, Shih-Fu Chang, and Arnold W. M. Smeulders (Eds.). ACM, 879–882. https://doi.org/10.1145/1873951.1874102

[62] Azhar Mohd Ibrahim, Ibrahim Venkat, K. G. Subramanian, Ahamad Tajudin Khader, and Philippe De Wilde. 2016. Intelligent Evacuation Management Systems: A Review. *ACM Trans. Intell. Syst. Technol.* 7, 3 (2016), 36:1–36:27. https://doi.org/10.1145/2842630

[63] Azhar Mohd Ibrahim, Ibrahim Venkat, and Philippe De Wilde. 2019. The Impact of Potential Crowd Behaviours on Emergency Evacuation: An Evolutionary Game-Theoretic Approach. *J. Artif. Soc. Soc. Simul.* 22, 1 (2019). https://doi.org/10.18564/jasss.3837

[64] Md Nafee Al Islam, Muhammed Tawfiq Chowdhury, Ankit Agrawal, Michael Murphy, Raj Mehta, Daria Kudriavtseva, Jane Cleland-Huang, Michael Vierhauser, and Marsha Chechik. 2023. Configuring mission-specific behavior in a product line of collaborating Small Unmanned Aerial Systems. *J. Syst. Softw.* 197 (2023), 111543. https://doi.org/10.1016/j.jss.2022.111543

[65] Kenneth Johnson, Javier Cámara, Roopak Sinha, Samaneh Madanian, and Dave Parry. 2021. Towards Self-Adaptive Disaster Management Systems. In *18th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2021, Blacksburg, VA, USA, May 2021*, Anouck Adrot, Rob Grace, Kathleen A. Moore, and Christopher W. Zobel (Eds.). ISCRAM Digital Library, 49–61. https://idl.iscram.org/show.php?record=2312

[66] Reshmashree B. Kantharaju and Catherine Pelachaud. 2021. Social Signals of Cohesion in Multi-party Interactions. In *IVA '21: ACM International Conference on Intelligent Virtual Agents, Virtual Event, Japan, September 14-17, 2021.* ACM, 9–16. https://doi.org/10.1145/3472306.3478362

[67] James Kendra, Jack Rozdilsky, and David A McEntire. 2008. Evacuating large urban areas: Challenges for emergency management policies and concepts. *Journal of Homeland Security and Emergency Management* 5, 1 (2008).

[68] Jeffrey O. Kephart and David M. Chess. 2003. The Vision of Autonomic Computing. *Computer* 36, 1 (2003), 41–50. https://doi.org/10.1109/MC.2003.1160055

[69] Nathan P. Koenig and Andrew Howard. 2004. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, September 28 - October 2, 2004.* IEEE, 2149–2154. https://doi.org/10.1109/IROS.2004.1389727

[70] Anastasia Kordoni, Carlos Gavidia-Calderon, Mark Levine, Amel Bennaceur, and Bashar Nuseibeh. 2023. "Are we in this together?": embedding social identity detection in drones improves emergency coordination. *Frontiers in Psychology* 14. https://doi.org/10.3389/fpsyg.2023.1146056

[71] Hadas Kress-Gazit, Kerstin Eder, Guy Hoffman, Henny Admoni, Brenna Argall, Rüdiger Ehlers, Christoffer Heckman, Nils Jansen, Ross A. Knepper, Jan Kretínský, Shelly Levy-Tzedek, Jamy Li, Todd D. Murphey, Laurel D. Riek, and Dorsa Sadigh. 2021. Formalizing and guaranteeing human-robot interaction. *Commun. ACM* 64, 9 (2021), 78–84. https://doi.org/10.1145/3433637

[72] William H. Kruskal and W. Allen Wallis. 1952. Use of Ranks in One-Criterion Variance Analysis. *J. Amer. Statist. Assoc.* 47, 260 (1952), 583–621. https://doi.org/10.1080/01621459.1952.10483441 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1952.10483441

[73] Max Kuhn, Kathrin Pollmann, and Judith Papadopoulos. 2020. I'm Your Partner - I'm Your Boss: Framing Human-Robot Collaboration with Conceptual Metaphors. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, HRI 2020, Cambridge, UK, March 23-26, 2020*, Tony Belpaeme, James E. Young, Hatice Gunes, and Laurel D. Riek (Eds.). ACM, 322–324. https://doi.org/10.1145/3371382.3378250

[74] Benjamin Kuipers. 2018. How can we trust a robot? *Commun. ACM* 61, 3 (2018), 86–95. https://doi.org/10.1145/3173087

[75] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. 2017. A unified game-theoretic approach to multiagent reinforcement learning. *Advances in neural information processing systems* 30 (2017).

[76] Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behrooz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey A. Voinov, J. Gary Polhill, Zhanli Sun, and Dawn Cassandra Parker. 2015. The Complexities of Agent-Based Modeling Output Analysis. *J. Artif. Soc. Soc. Simul.* 18, 4 (2015). https://doi.org/10.18564/jasss.2897

[77] Mark Levine and Clare Cassidy. 2009. *Groups, Identities, and Bystander Behavior.* John Wiley & Sons, Ltd, Chapter 11, 209–222. https://doi.org/10.1002/9781444307948.ch11 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781444307948.ch11

[78] Mark Levine, Amy Prosser, David Evans, and Stephen Reicher. 2005. Identity and Emergency Intervention: How Social Group Membership and Inclusiveness of Group Boundaries Shape Helping Behavior. *Personality and Social Psychology Bulletin* 31, 4 (2005), 443–453. https://doi.org/10.1177/0146167204271651 arXiv:https://doi.org/10.1177/0146167204271651 PMID: 15743980.

[79] Kevin Leyton-Brown and Yoav Shoham. 2008. *Essentials of Game Theory: A Concise Multidisciplinary Introduction.* Morgan & Claypool Publishers. https://doi.org/10.2200/S00108ED1V01Y200802AIM003

[80] Nianyu Li, Sridhar Adepu, Eunsuk Kang, and David Garlan. 2020. Explanations for human-on-the-loop: a probabilistic model checking approach. In *SEAMS '20: IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, Seoul, Republic of Korea, 29 June - 3 July, 2020*, Shinichi Honiden, Elisabetta Di Nitto, and Radu Calinescu (Eds.). ACM, 181–187. https://doi.org/10.1145/3387939.3391592

[81] Nianyu Li, Javier Cámara, David Garlan, Bradley R. Schmerl, and Zhi Jin. 2021. Hey! Preparing Humans to do Tasks in Self-adaptive Systems. In *16th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS@ICSE 2021, Madrid, Spain, May 18-24, 2021.* IEEE, 48–58. https://doi.org/10.1109/SEAMS51251.2021.00017

[82] Yanan Li, Gerolamo Carboni, Franck Gonzalez, Domenico Campolo, and Etienne Burdet. 2019. Differential game theory for versatile physical human-robot interaction. *Nat. Mach. Intell.* 1, 1 (2019), 36–43. https://doi.org/10.1038/s42256-018-0010-3

[83] Hong Liu, Bin Xu, Dianjie Lu, and Guijuan Zhang. 2018. A path planning approach for crowd evacuation in buildings based on improved artificial bee colony algorithm. *Appl. Soft Comput.* 68 (2018), 360–376. https://doi.org/10.1016/j.asoc.2018.04.015

[84] David Lo. 2023. Trustworthy and Synergistic Artificial Intelligence for Software Engineering: Vision and Roadmaps. *CoRR* abs/2309.04142 (2023). https://doi.org/10.48550/ARXIV.2309.04142 arXiv:2309.04142

[85] Jabier Martinez, Wesley KG Assunção, and Tewfik Ziadi. 2017. ESPLA: A catalog of Extractive SPL Adoption case studies. In *21st International Systems and Software Product Line Conference.* ACM, 38–41.

[86] Richard D. McKelvey, Andrew M.and McLennan, and Theodore L. Turocy. 2021. Gambit: Software Tools for Game Theory, Version 15. http://www.gambit-project.org

[87] Alberto Melucci et al. 1995. The process of collective identity. *Social movements and culture* 4 (1995), 41–63.

[88] Costanza Messeri, Gabriele Masotti, Andrea Maria Zanchettin, and Paolo Rocco. 2021. Human-Robot Collaboration: Optimizing Stress and Productivity Based on Game Theory. *IEEE Robotics Autom. Lett.* 6, 4 (2021), 8061–8068. https://doi.org/10.1109/LRA.2021.3102309

[89] Johannes Meyer, Paul Schnitzspan, Stefan Kohlbrecher, Karen Petersen, Mykhaylo Andriluka, Oliver Schwahn, Uwe Klingauf, Stefan Roth, Bernt Schiele, and Oskar von Stryk. 2010. A Semantic World Model for Urban Search and Rescue Based on Heterogeneous Sensors. In *RoboCup 2010: Robot Soccer World Cup XIV [papers from the 14th annual RoboCup International Symposium, Singapore, June 25, 2010] (Lecture Notes in Computer Science, Vol. 6556)*, Javier Ruiz-del-Solar, Eric Chown, and Paul-Gerhard Plöger (Eds.). Springer, 180–193. https://doi.org/10.1007/978-3-642-20217-9_16

[90] Henry Muccini, Mohammad Sharaf, and Danny Weyns. 2016. Self-adaptation for cyber-physical systems: a systematic literature review. In *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS@ICSE 2016, Austin, Texas, USA, May 14-22, 2016.* ACM, 75–81. https://doi.org/10.1145/2897053.2897069

[91] R.R. Murphy, D. Riddle, and E. Rasmussen. 2004. Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759).* 301–306. https://doi.org/10.1109/ROMAN.2004.1374777

[92] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani (Eds.). 2007. *Algorithmic Game Theory.* Cambridge University Press. https://doi.org/10.1017/CBO9780511800481

[93] Ignacio Palacios-Huerta. 2003. Professionals Play Minimax. *The Review of Economic Studies* 70, 2 (04 2003), 395–415. https://doi.org/10.1111/1467-937X.00249 arXiv:https://academic.oup.com/restud/article-pdf/70/2/395/4504659/70-2-395.pdf

[94] Richard Philpot and Mark Levine. 2022. Evacuation behavior in a subway train emergency: a video-based analysis. *Environment and Behavior* 54, 2 (2022), 383–411.

[95] David Pietraszewski. 2022. Toward a computational theory of social groups: A finite set of cognitive primitives for representing any and all social groups in the context of conflict. *Behavioral and Brain Sciences* 45 (2022), e97. https://doi.org/10.1017/S0140525X21000583

[96] Nicola Power and Laurence Alison. 2017. Offence or defence? Approach and avoid goals in the multi-agency emergency response to a simulated terrorism attack. *Journal of occupational and organizational psychology* 90, 1 (2017), 51–76.

[97] Nicola Power and Laurence Alison. 2017. Redundant deliberation about negative consequences: Decision inertia in emergency responders. *Psychology, public policy, and law* 23, 2 (2017), 243.

[98] Anna Priante, Michel L Ehrenhard, Tijs van den Broek, and Ariana Need. 2018. Identity and collective action via computer-mediated communication: A review and agenda for future research. *New Media & Society* 20, 7 (2018), 2647–2669. https://doi.org/10.1177/1461444817744783 arXiv:https://doi.org/10.1177/1461444817744783 PMID: 30581365.

[99] Salil Purandare, Urjoshi Sinha, Md Nafee Al Islam, Jane Cleland-Huang, and Myra B. Cohen. 2023. Self-Adaptive Mechanisms for Misconfigurations in Small Uncrewed Aerial Systems. In *18th IEEE/ACM Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2023, Melbourne, Australia, May 15-16, 2023*. IEEE, 169–180. https://doi.org/10.1109/SEAMS59076.2023.00030

[100] Jorge Pena Queralta, Jussi Taipalmaa, Bilge Can Pullinen, Victor Kathan Sarker, Tuan Nguyen Gia, Hannu Tenhunen, Moncef Gabbouj, Jenni Raitoharju, and Tomi Westerlund. 2020. Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision. *Ieee Access* 8 (2020), 191617–191643.

[101] Itzhak Rasooly and Carlos Gavidia-Calderon. 2020. The importance of being discrete: on the inaccuracy of continuous approximations in auction theory. *arXiv preprint arXiv:2006.03016* (2020).

[102] Irum Rauf, Dirk van der Linden, Mark Levine, John N. Towse, Bashar Nuseibeh, and Awais Rashid. 2020. Security but not for security's sake: The impact of social considerations on app developers' choices. In *ICSE '20: 42nd International Conference on Software Engineering, Workshops, Seoul, Republic of Korea, 27 June - 19 July, 2020*. ACM, 141–144. https://doi.org/10.1145/3387940.3392230

[103] Mark N. Read, Kieran Alden, Jon Timmis, and Paul S. Andrews. 2020. Strategies for calibrating models of biology. *Briefings Bioinform.* 21, 1 (2020), 24–35. https://doi.org/10.1093/bib/bby092

[104] Stephen Reicher. 2015. Mass action and mundane reality: An argument for putting crowd analysis at the centre of the social sciences. In *Crowds in the 21st Century*. Routledge, 169–185.

[105] Stephen Reicher and Russell Spears. 2010. The Social Identity Approach in Social Psychology. *The SAGE Handbook of Identities* (2010), 45.

[106] George Ritter, Robert A Wolfe, Sidney Goldstein, J Richard Landis, C Mark Vasu, Allyn Acheson, Richard Leighton, and Sharon Vander-Brug Medendrop. 1985. The effect of bystander CPR on survival of out-of-hospital cardiac arrest victims. *American heart journal* 110, 5 (1985), 932–937.

[107] Giacomo Rizzolatti and Corrado Sinigaglia. 2008. *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press, USA.

[108] Simcha Ronen and Oded Shenkar. 2013. Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies* 44, 9 (2013), 867–897. https://EconPapers.repec.org/RePEc:pal:jintbs:v:44:y:2013:i:9:p:867-897

[109] Tuomas Sandholm. 2015. Abstraction for Solving Large Incomplete-Information Games. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, Blai Bonet and Sven Koenig (Eds.). AAAI Press, 4127–4131. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/10039

[110] Viktor Schmuck and Oya Çeliktutan. 2021. GROWL: Group Detection With Link Prediction. In *16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, December 15-18, 2021*. IEEE, 1–8. https://doi.org/10.1109/FG52635.2021.9667061

[111] Eike Schneiders. 2022. Non-Dyadic Collaboration In Human-Robot Interaction. (2022).

[112] Eike Schneiders, EunJeong Cheon, Jesper Kjeldskov, Matthias Rehm, and Mikael B. Skov. 2022. Non-Dyadic Interaction: A Literature Review of 15 Years of Human-Robot Interaction Conference Publications. *J. Hum.-Robot Interact.* 11, 2, Article 13 (feb 2022), 32 pages. https://doi.org/10.1145/3488242

[113] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. 2019. Social behavior for autonomous vehicles. *Proc. Natl. Acad. Sci. USA* 116, 50 (2019), 24972–24978. https://doi.org/10.1073/pnas.1820676116

[114] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. 2006. Joint action: bodies and minds moving together. *Trends in cognitive sciences* 10, 2 (2006), 70–76.

[115] Joseph Seering, Felicia Ng, Zheng Yao, and Geoff Kaufman. 2018. Applications of Social Identity Theory to Research and Design in Computer-Supported Cooperative Work. *Proc. ACM Hum. Comput. Interact.* 2, CSCW (2018), 201:1–201:34. https://doi.org/10.1145/3274771

[116] Mary Shaw. 2002. What makes good research in software engineering? *International Journal on Software Tools for Technology Transfer* 4, 1 (2002), 1–7.

[117] Dhirendra Singh and Lin Padgham. 2015. Community Evacuation Planning for Bushfires Using Agent-Based Simulation: Demonstration. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, Gerhard Weiss, Pinar Yolum, Rafael H. Bordini, and Edith Elkind (Eds.). ACM, 1903–1904. http://dl.acm.org/citation.cfm?id=2773496

[118] Dhirendra Singh and Lin Padgham. 2017. Emergency Evacuation Simulator (EES) - a Tool for Planning Community Evacuations in Australia. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, Carles Sierra (Ed.). ijcai.org, 5249–5251. https://doi.org/10.24963/ijcai.2017/780

[119] Paul E. Smaldino, Cynthia Pickett, Jeffrey Sherman, and Jeffrey C. Schank. 2012. An Agent-Based Model of Social Identity Dynamics. *J. Artif. Soc. Soc. Simul.* 15, 4 (2012). https://doi.org/10.18564/jasss.2030

[120] Russell Spears. 2021. Social influence and group identity. *Annual Review of Psychology* 72 (2021), 367–390.

[121] Paul W Speer and N Andrew Peterson. 2000. Psychometric properties of an empowerment scale: Testing cognitive, emotional, and behavioral domains. *Social work research* 24, 2 (2000), 109–118.

[122] Avelie Stuart, Dmitri Katz, Clifford Stevenson, Daniel Gooch, Lydia Harkin, Mohamed Bennasar, Lisa Sanderson, Jacki Liddle, Amel Bennaceur, Mark Levine, et al. 2022. Loneliness in older people and COVID-19: applying the social identity approach to digital intervention design. *Computers in Human Behavior Reports* (2022), 100179.

[123] William B Swann Jr, Angel Gómez, D Conor Seyle, J Morales, and Carmen Huici. 2009. Identity fusion: the interplay of personal and social identities in extreme group behavior. *Journal of personality and social psychology* 96, 5 (2009), 995.

[124] Henri Tajfel. 1982. Social psychology of intergroup relations. *Annual review of psychology* 33, 1 (1982), 1–39.

[125] Henri Tajfel and John C Turner. 1979. An Integrative Theory of Intergroup Conflict. *The Social Psychology of Intergroup Relations* 33, 47 (1979), 74.

[126] Yunus Terzioglu, Bilge Mutlu, and Erol Sahin. 2020. Designing Social Cues for Collaborative Robots: The Role of Gaze and Breathing in Human-Robot Collaboration. In *HRI '20: ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, United Kingdom, March 23-26, 2020*, Tony Belpaeme, James E. Young, Hatice Gunes, and Laurel D. Riek (Eds.). ACM, 343–357. https://doi.org/10.1145/3319502.3374829

[127] Lewis Tunstall, Leandro von Werra, and Thomas Wolf. 2022. *Natural Language Processing with Transformers: Building Language Applications with Hugging Face*. O'Reilly Media, Incorporated. https://books.google.ch/books?id=7hhyzgEACAAJ

[128] John C Turner, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, Cambridge, MA, US. x, 239–x, 239 pages.

[129] Sara M. Vambheim, Silje C. Wangberg, Jan-Are K. Johnsen, and Rolf Wynn. 2013. Language use in an internet support group for smoking cessation: development of sense of community. *Informatics for Health and Social Care* 38, 1 (2013), 67–78. https://doi.org/10.3109/17538157.2012.710685 arXiv:https://doi.org/10.3109/17538157.2012.710685 PMID: 22958060.

[130] C. Natalie van der Wal, Daniel Formolo, Mark A. Robinson, and Steven Gwynne. 2021. Examining Evacuee Response to Emergency Communications with Agent-Based Simulations. *Sustainability* 13, 9 (2021). https://doi.org/10.3390/su13094623

[131] C. Natalie van der Wal, Daniel Formolo, Mark A. Robinson, Michael Minkov, and Tibor Bosse. 2017. Simulating Crowd Evacuation with Socio-Cultural, Cognitive, and Emotional Elements. *Trans. Comput. Collect. Intell.* 27 (2017), 139–177. https://doi.org/10.1007/978-3-319-70647-4_11

[132] Christopher Van Tilburg. 2017. First report of using portable unmanned aircraft systems (drones) for search and rescue. *Wilderness & environmental medicine* 28, 2 (2017), 116–118.

[133] Eugene Vinitsky, Raphael Köster, John P. Agapiou, Edgar A. Duéñez-Guzmán, Alexander Sasha Vezhnevets, and Joel Z. Leibo. 2021. A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. *CoRR* abs/2106.09012 (2021). arXiv:2106.09012 https://arxiv.org/abs/2106.09012

[134] Birgit Vogel-Heuser, Stefan Feldmann, Jens Folmer, Jan Ladiges, Alexander Fay, Sascha Lity, Matthias Tichy, Matthias Kowal, Ina Schaefer, Christopher Haubeck, et al. 2015. Selected challenges of software evolution for automated production systems. In *13th International Conference on Industrial Informatics*. IEEE, 314–321.

[135] Isabella von Sivers, Anne Templeton, Florian Künzner, Gerta Köster, John Drury, Andrew Philippides, Tobias Neckel, and H-J Bungartz. 2016. Modelling social identification and helping in evacuation simulation. *Safety science* 89 (2016), 288–300.

[136] Alan R. Wagner, Paul Robinette, and Ayanna M. Howard. 2018. Modeling the Human-Robot Trust Phenomenon: A Conceptual Framework based on Risk. *ACM Trans. Interact. Intell. Syst.* 8, 4 (2018), 26:1–26:24. https://doi.org/10.1145/3152890

[137] Mark Walker and John Wooders. 2001. Minimax Play at Wimbledon. *American Economic Review* 91, 5 (December 2001), 1521–1538. https://doi.org/10.1257/aer.91.5.1521

[138] Danny Weyns. 2019. Software Engineering of Self-adaptive Systems. In *Handbook of Software Engineering*, Sungdeok Cha, Richard N. Taylor, and Kyo Chul Kang (Eds.). Springer, 399–443. https://doi.org/10.1007/978-3-030-00262-6_11

[139] Danny Weyns, Radu Calinescu, Raffaela Mirandola, Kenji Tei, Maribel Acosta, Amel Bennaceur, Nicolas Boltz, Tomas Bures, Javier Camara, Ada Diaconescu, Gregor Engels, Simos Gerasimou, Ilias Gerostathopoulos, Sinem Getir Yaman, Vincenzo Grassi, Sebastian Hahner, Emmanuel Letier, Marin Litoiu, Lina Marsso, Angelika Musil, Juergen Musil, Genaina Nunes Rodrigues, Diego Perez-Palacin, Federico Quin, Patrizia Scandurra, Antonio Vallecillo, and Andrea Zisman. 2023. Towards a Research Agenda for Understanding and ManagingUncertainty in Self-Adaptive Systems. *SIGSOFT Softw. Eng. Notes* 48, 4 (oct 2023), 20–36. https://doi.org/10.1145/3617946.3617951

[140] Danny Weyns, Bradley R. Schmerl, Vincenzo Grassi, Sam Malek, Raffaela Mirandola, Christian Prehofer, Jochen Wuttke, Jesper Andersson, Holger Giese, and Karl M. Göschka. 2010. On Patterns for Decentralized Control in Self-Adaptive Systems. In *Software Engineering for Self-Adaptive Systems II - International Seminar, Dagstuhl Castle, Germany, October 24-29, 2010 Revised Selected and Invited Papers (Lecture Notes in Computer Science, Vol. 7475)*, Rogério de Lemos, Holger Giese, Hausi A. Müller, and Mary Shaw (Eds.). Springer, 76–107. https://doi.org/10.1007/978-3-642-35813-5_4

[141] Nanda Wijermans and Anne Templeton. 2022. Towards more realism in pedestrian behaviour models: First steps and considerations in formalising social identity. In *Advances in Social Simulation: Proceedings of the 16th Social Simulation Conference, 20–24 September 2021*. Springer, 53–64.

[142] Ronald T Wilcox. 2000. Experts and amateurs: The role of experience in Internet auctions. *Marketing Letters* 11, 4 (2000), 363–374.

[143] Uri Wilensky. 1999. NetLogo. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL* (1999). https://ccl.northwestern.edu/netlogo/

[144] Richard Williams and John Drury. 2009. Psychosocial resilience and its influence on managing mass emergencies and disasters. *Psychiatry* 8, 8 (2009), 293–296. https://doi.org/10.1016/j.mppsy.2009.04.019 Trauma & stress-related disorder.

[145] Alexandra Willis, Nathalia Gjersoe, Catriona Havard, Jon Kerridge, and Robert Kukla. 2004. Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments. *Environment and Planning B: Planning and Design* 31, 6 (2004), 805–828.

[146] Michael J. Wooldridge. 2012. Does Game Theory Work? *IEEE Intell. Syst.* 27, 6 (2012), 76–80. https://doi.org/10.1109/MIS.2012.108

[147] James R. Wright and Kevin Leyton-Brown. 2012. Behavioral game theoretic models: a Bayesian framework for parameter analysis. In *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4-8, 2012 (3 Volumes)*, Wiebe van der Hoek, Lin Padgham, Vincent Conitzer, and Michael Winikoff (Eds.). IFAAMAS, 921–930. http://dl.acm.org/citation.cfm?id=2343828

[148] Jane Wu, Erin Paeng, Kari Linder, Piercarlo Valdesolo, and James C. Boerkoel. 2016. Trust and Cooperation in Human-Robot Decision Making. In *2016 AAAI Fall Symposia, Arlington, Virginia, USA, November 17-19, 2016*. AAAI Press. http://aaai.org/ocs/index.php/FSS/FSS16/paper/view/14118

[149] Xiaoping Zheng, Tingkuan Zhong, and Mengting Liu. 2009. Modeling crowd evacuation of a building based on seven methodological approaches. *Building and Environment* 44, 3 (2009), 437–445. https://doi.org/10.1016/j.buildenv.2008.04.002

[150] Kashif Zia and Alois Ferscha. 2020. An Agent-Based Model of Crowd Evacuation: Combining Individual, Social and Technological Aspects. In *Proceedings of the 2019 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, SIGSIM-PADS 2020, Miami, FL, USA, June 15-17, 2020*, Jason Liu, Philippe J. Giabbanelli, and Christopher D. Carothers (Eds.). ACM, 129–140. https://doi.org/10.1145/3384441.3395973

[151] Shlomo Zilberstein. 2015. Building Strong Semi-Autonomous Systems. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, Blai Bonet and Sven Koenig (Eds.). AAAI Press, 4088–4092. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9920