



## Engineering multiuser museum interactives for shared cultural experiences



Roberto Confalonieri<sup>a,\*</sup>, Matthew Yee-King<sup>b</sup>, Katina Hazelden<sup>b</sup>, Mark d'Inverno<sup>b</sup>,  
Dave de Jonge<sup>a</sup>, Nardine Osman<sup>a</sup>, Carles Sierra<sup>a</sup>, Leila Agmoud<sup>c</sup>, Henri Prade<sup>c</sup>

<sup>a</sup> Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain

<sup>b</sup> Goldsmiths College, University of London, London, UK

<sup>c</sup> Institute de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier, Toulouse, France

### ARTICLE INFO

#### Article history:

Received 27 March 2015

Received in revised form

16 July 2015

Accepted 28 August 2015

#### Keywords:

Distributed artificial intelligence

Collective decision making

Argumentation

Negotiation

### ABSTRACT

Multiuser museum interactives are computer systems installed in museums or galleries which allow several visitors to interact together with digital representations of artefacts and information from the museum's collection. In this paper, we describe WeCurate, a socio-technical system that supports co-browsing across multiple devices and enables groups of users to collaboratively curate a collection of images, through negotiation, collective decision making and voting. The engineering of such a system is challenging since it requires to address several problems such as: distributed workflow control, collective decision making and multiuser synchronous interactions. The system uses a peer-to-peer Electronic Institution (EI) to manage and execute a distributed curation workflow and models community interactions into scenes, where users engage in different social activities. Social interactions are enacted by intelligent agents that interface the users participating in the curation workflow with the EI infrastructure. The multiagent system supports collective decision making, representing the actions of the users within the EI, where the agents advocate and support the desires of their users e.g. aggregating opinions for deciding which images are interesting enough to be discussed, and proposing interactions and resolutions between disagreeing group members. Throughout the paper, we describe the enabling technologies of WeCurate, the peer-to-peer EI infrastructure, the agent collective decision making capabilities and the multi-modal interface. We present a system evaluation based on data collected from cultural exhibitions in which WeCurate was used as supporting multiuser interactive.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

In recent times, high tech museum interactives have become ubiquitous in major institutions. Typical examples include augmented reality systems, multitouch table tops and virtual reality tours (Gaitatzes and Roussou, 2002; Hornecker, 2008; Wojciechowski et al., 2004). Whilst multiuser systems have begun to appear, e.g. a 10 user quiz game in the Tate Modern, the majority of these museum interactives do not perhaps facilitate the socio-cultural experience of visiting a museum with friends, as they are often being designed for a single user. The need to support

multiuser interaction and social participation is a desirable feature for shifting the focus from content delivery to social construction (Walker, 2008) and for the development of a cultural capital (Hope et al., 2009).

At this point, we should note that mediating and reporting the actions of several 'agents' to provide a meaningful and satisfying sociocultural experience for all is challenging (Heath et al., 2005). Social interaction and collaboration are key features for the development of a socio-technical system like the one described in this paper. On the one hand, the system has to enhance user interactions and should be accessible independently from user locations. This requires a robust and flexible infrastructure that is able to capture a *social workflow* and the *dynamics* of the community which will engage in the system. On the other hand, the system has to assist users in *collective decision making* and *negotiation*, and to foster participation and discussions about the cultural artefacts. This requires the use of autonomic agents that can advocate and support the desires of their users e.g. aggregating

\* Corresponding author.

E-mail addresses: [confalonieri@iiia.csic.es](mailto:confalonieri@iiia.csic.es) (R. Confalonieri), [m.yee-king@gold.ac.uk](mailto:m.yee-king@gold.ac.uk) (M. Yee-King), [kat9@me.com](mailto:kat9@me.com) (K. Hazelden), [dinverno@gold.ac.uk](mailto:dinverno@gold.ac.uk) (M. d'Inverno), [davedejonge@iiia.csic.es](mailto:davedejonge@iiia.csic.es) (D. de Jonge), [nardine@iiia.csic.es](mailto:nardine@iiia.csic.es) (N. Osman), [sierra@iiia.csic.es](mailto:sierra@iiia.csic.es) (C. Sierra), [agmoud@irit.fr](mailto:agmoud@irit.fr) (L. Agmoud), [prade@irit.fr](mailto:prade@irit.fr) (H. Prade).

opinions for deciding which images are interesting enough to be discussed, and proposing interactions and resolutions between disagreeing group members.

Another trend in museum curation is the idea of community curation, where a community discourse is built up around the artefacts, to provide different perspectives and insights (Turner, 2011). This trend is not typically represented in the design of museum interactives, where information-browsing, and not information-generation is the focus. However, museums are engaging with the idea of crowdsourcing, with projects such as “Your Paintings Tagger” and “The Art Of Video Games” (Greg, 2011; Barron, 2012), and folksonomies with projects such as “steve.project” and “Artlinks” (Hellin-Hobbs, 2010; Cosley and Lewenstein, 2008; Cosley and Baxter, 2009). Again, controlling the workflow within a group to engender discussion and engagement with the artefacts is challenging, especially when the users are casual ones as in a museum context.

In this paper, we describe WeCurate, a first of its kind multiuser museum interactive. WeCurate uses a multiagent system to support community interactions and decision making, and a peer-to-peer Electronic Institution (EI) (de Jonge et al., 2013) to execute and control the community workflow. Our aim is not only to make use of agent technology and Electronic Institutions as a means to implement a multiuser museum interactive, but also to relate agent theory to practice in order to create a socio-technical system to support an online multiuser experience.

To this end, we specify a community curation session in terms of the scenes of an EI for controlling community interactions. We support system and user decisions by means of personal assistant agents equipped with different decision making capabilities. We make use of a multimodal user interface which directly represents users as agents in the scenes of the underlying EI and which is designed to engage casual users in a social discourse around museum artefacts by chat and tag activity. We present the evaluation of the system for determining the level of interactions and social awareness perceived by the social groups while using the system, and for understanding whether our agent-based decision models can predict what images users like from their behaviour. We validate our scene-based design and, consequently, our EI model, from the social behaviour of users that emerged naturally during the curation task.

This paper unifies and develops the content of the conference papers (Amgoud et al., 2012; Yee-King et al., 2013; Hazelden et al., 2013) by describing the underlying peer-to-peer EI infrastructure and presenting an analysis of the decision making models employed by the agents. The evaluation is based on data collected from cultural exhibitions in which WeCurate was used as a supporting multiuser museum interactive. The rest of the paper is organised as follows. Section 2 provides an overview of the system, whereas Section 3, 4, 5, and 6 respectively describe the EI infrastructure and workflow, the personal assistant agents, the interface and the adopted technologies. Section 7 presents the evaluation of our system. After discussing the evaluation's results (Section 8), Section 9 presents several works that relate to ours from different perspectives. Finally, in Section 10 we draw some conclusions and we envision some of the ideas we have in mind to improve the current system.

## 2. System overview

WeCurate is a museum interactive which provides a multiuser curation workflow where the aim is for the users to synchronously view and discuss a selection of images, finally choosing a subset of these images that the group would like to add to their group collection. In the process of curating this collection, the users are encouraged to develop a discourse about the images in the form of

weighted tags and comments, as well as a process of bilateral argumentation. Further insight into user preferences and behaviours is gained from data about specific user actions such as image zooming and general activity levels.

A multiuser interactive is a typical example of a system in which human and software agents can enter and leave the system and behave according to the *norms* that are appropriate for that specific society. For instance, it can be desirable to have only a certain number of users taking part to a curation session or to allow each user to express at most one vote. A convenient way to coordinate the social interactions of agent communities is by means of an Electronic Institution (EI) (Arcos et al., 2005).

An EI makes it possible to develop programs according to a new paradigm, in which the tasks are executed by independent agents, that are not specifically designed for the given program and that cannot be blindly trusted. An EI is responsible for making sure that the agents behave according to the norms that are necessary for the application. To this end, the actions that agents can perform in an EI are represented as messages and are specified according to an interaction protocol for each scene. The EI checks for each message whether it is valid in the current state of the protocol, and, if not, prevents it from being delivered to the other agents participating in the EI. In this way, the behaviour of non-benevolent agents can be controlled.<sup>1</sup> Therefore, the EI paradigm allows a flexible and dynamic infrastructure, in which agents can interact in an autonomous way within the norms of the cultural institution.

EIs have usually been considered as centralised systems (Noriega, 1997; Esteva, 2003). Nevertheless, the growing need to incorporate organisational abstractions into distributed computing systems (d'Inverno et al., 2012) requires a new form of EIs.

In WeCurate, since users can be physically in different places, it is desirable to run an EI in a distributed manner to characterise human social communities in a more natural manner. To this end, we implemented a new form of EI that runs in a distributed way, over a peer-to-peer network (de Jonge et al., 2013). The multiuser curation workflow has been modeled as scenes of an EI and scene protocols. The workflow is managed and executed by a peer-to-peer EI, with agents operating within it to represent the activities of the users and to provide other services. The users interact with the system using an animated user interface. An overview of the system architecture, showing the peer-to-peer EI, the User Assistant agents and user interface components are provided in Fig. 1.

In the following sections, we present the internal structure of the peer-to-peer Electronic Institution and the WeCurate curation workflow. Then, we describe the agents that participate in the workflow, with particular emphasis on user representation and collective decision making. The user interface is presented with images of the different scenes in the workflow. The system architecture is described, including the connections between EI, agents and UI. Finally, the adopted technologies used to implement the system are briefly explained.

<sup>1</sup> The EI cannot control, however, the behaviour of a non-benevolent agent when it fails to perform an action that the protocol requires it to perform. It essentially cannot force an agent to do something it does not wish to do. This is because EIs are designed for autonomous agents, and although we would like agents to behave in certain ways, their autonomy must be maintained. In such a case, either the protocol engineer can make use of timeouts to make the protocols resilient against such scenarios, or misbehaviour should be addressed through other measures, such as sanctions and rewards (Modgil et al., 2009; Gaertner et al., 2007), trust and reputation (Osman et al., 2014), and so on.

The EI also cannot control the behaviour of a non-benevolent agent that *does* follow a protocol but does it in a malicious way, for instance, by pretending to like an image, or by pushing other users to change their opinion with no specific reason, etc. To address this situation, again trust models can be used to detect and block the malicious behaviour of an agent, for instance, by assessing the trustworthiness of an agent through learning from similar past experiences (Osman et al., 2014).

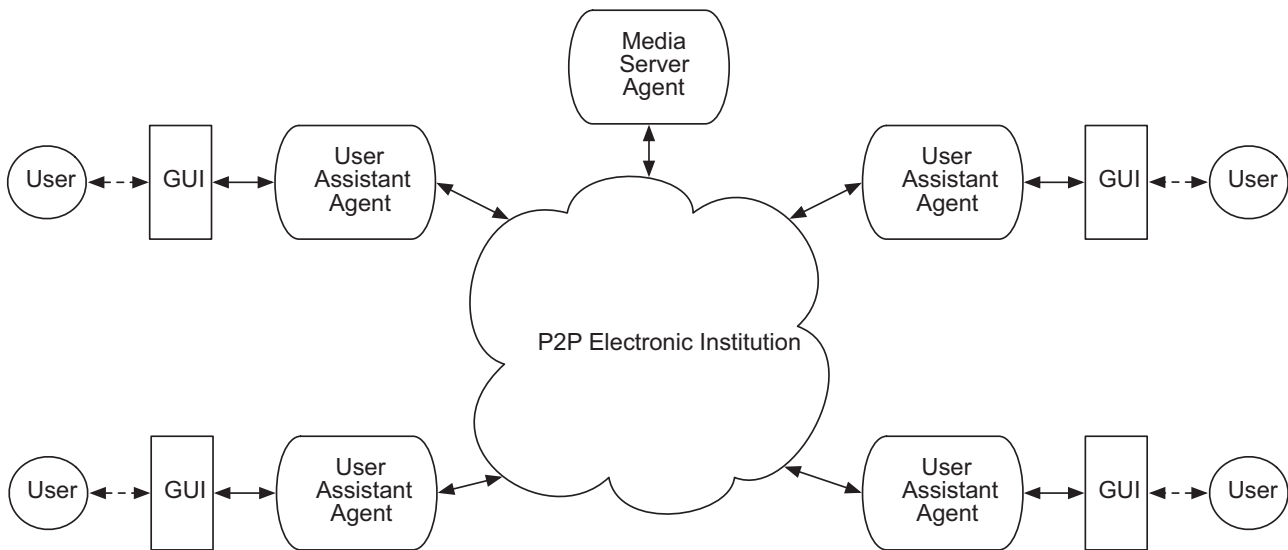


Fig. 1. The WeCurate system architecture, here showing 4 simultaneous users.

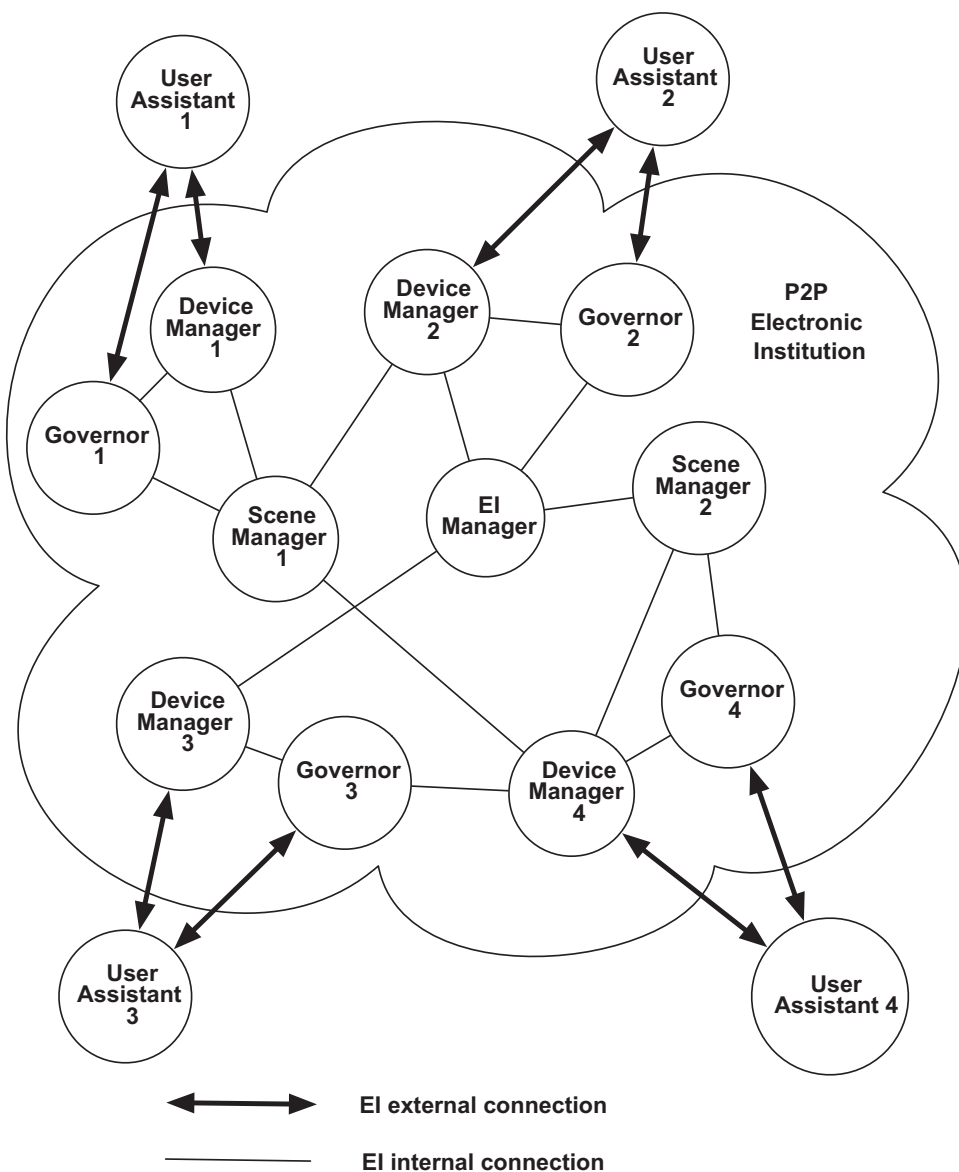


Fig. 2. Structure of the p2p electronic institution. Note that the external agents do *not* form part of the p2p-network. The connections in this diagram are drawn randomly.

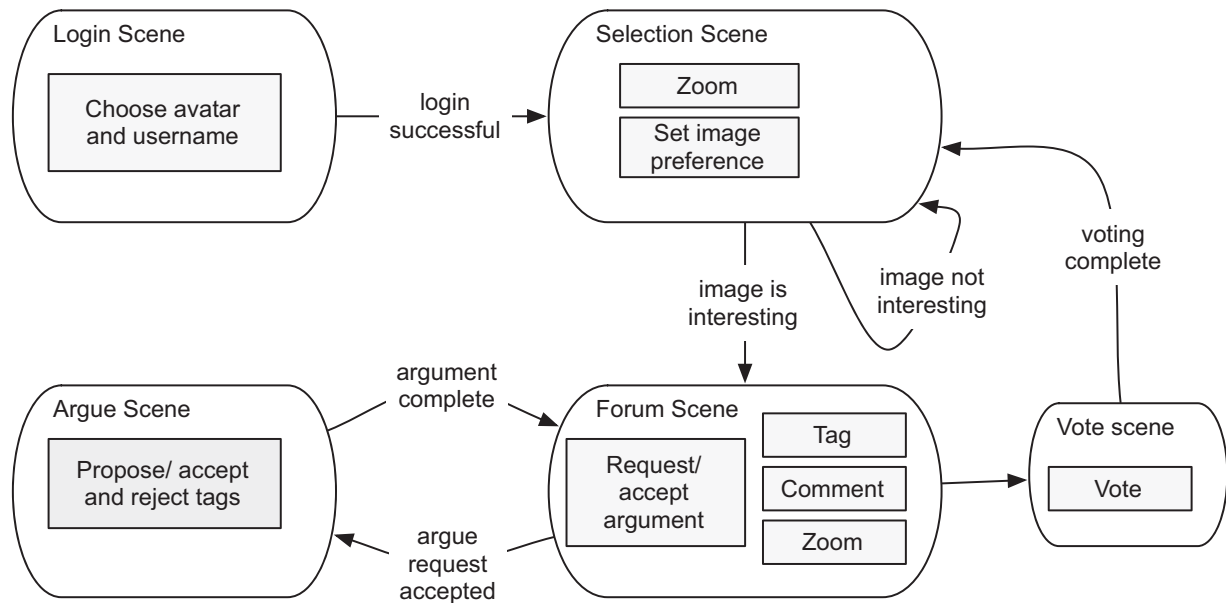


Fig. 3. The WeCurate workflow: white boxes represent scenes, grey boxes represent user actions, and arrows denote scene transitions.

### 3. Peer-to-peer Electronic Institution

The structure of the peer-to-peer EI is displayed in Fig. 2. The EI itself is executed by several institutional agents, including a *Scene Manager* which runs the scene instances, an *EI Manager* which admits *External Agents* to the EI and instantiates scenes, and several *Governors* which control message passing between agents:

- *External Agent*: The term *External Agent* is a generic term that represents any type of agent that can participate in an EI. It should be distinguished from the other agents described below which are *Institutional Agents* and are responsible for making the EI operate properly. A *User Assistant* is a specific type of *External Agent* that acts as an interface between a human user and the EI. It allows users to 'enter' the institution. In some cases, an *External Agent* may just have an interface that passes messages from humans to EI and vice-versa, while in other cases it can have more functionalities such as an intelligent module to help users making decisions. As we shall see, an agent might assist the users in negotiations and bilateral argumentation sessions with other agents.
- *Governor*: The Governor is an agent assigned to each *External Agent* participating in the EI to control the *External Agent* behaviour. Governors form a protected layer between the external agents and the institution. Since each action an agent can take within the institution is represented by a message, the Governor performs its task by checking whether a message sent by the agent is allowed in the current context of the institution.
- *Device Manager*: The Device Manager is a component that we introduce specifically for the peer-to-peer EI. A Device Manager is in charge of launching the *Institutional Agents* on its local device, and, if necessary, requests other Device Managers on other devices to do so. The motivation for introducing Device Managers, is that in a mobile network the present devices usually have varying capabilities, often limited, and therefore one should find a suitable balance of work load between the devices. Moreover, since for most institutional agents it does not matter on what device they are running, we need a system to

determine where they will be launched. We assume that each device in the network has exactly one device manager. The Device Manager is not bound to one specific instance of the EI; it may run agents from several different institutions.

- *EI Manager*: The EI manager is the agent that is responsible for admitting agents into the institution and for instantiating and launching scenes.
- *Scene Manager*: Each scene instance is assigned a Scene Manager. The Scene Manager is responsible for making sure the scene functions properly. It records all context variables of the scene.

The peer-to-peer EI infrastructure described above manages distributed workflows modelled as EI specifications. An EI specification consists of scenes and scene protocols. Scenes are essentially 'meeting rooms' in which agents can meet and interact. Scene protocols are well-defined communication protocols that specify the possible dialogues between agents within these scenes. Scenes within an institution are connected in a network that determines how agents can legally move from one scene to another through scene transitions. The EI specification is then interpreted by a workflow engine which controls the workflow execution and the messages sent over the EI. We omit the details about the EI specification language and the EI workflow engine; the reader can find a more extensive description in de Jonge et al. (2013), Arcos et al. (2005) and de Jonge et al. (2014). In what follows, we present the workflow we used for modelling the activity of community curation carried out by the users in the WeCurate system, and how we implement scene transitions as decision making models of the agents.

#### 3.1. WeCurate workflow

The WeCurate workflow consists of 5 scenes, with associated rules controlling messaging and transitions between scenes. An overview of the workflow is provided in Fig. 3. The scenes are as follows:



**Fig. 4.** The WeCurate user interface. Bubbles represent tags and are resizable and movable; icons visible on sliders and images represent users. (a) The select scene for rapid selection of interesting images. (b) The forum scene for in-depth discussion of images. (c) The argue scene for bilateral argumentation among two users. (d) The vote scene for deciding to add images to the group collection.

- **Login and lobby scene:** This allows users to login and wait for other users to join. The EI can be configured to require a certain number of users to login before the transition to the selection scene can take place.
- **Selection scene:** Its purpose is to allow a quick decision as to whether an image is interesting enough for a full discussion. Users can *zoom* into the image and see the zooming actions of other users. They can also set their *overall preference* for the image using a like/dislike slider. The user interface of this scene is shown in Fig. 4(a).
- **Forum scene:** If an image is deemed interesting enough, the users are taken to the forum scene where they can engage in a *discussion* about the image. Users can add and delete tags, they can resize tags to define their opinions of that aspect of the image, they can make comments, they can zoom into the image and they can see the actions of the other users. They can also view images that were previously added to the collection and choose to argue with another user directly. The aim is to collect community information about the image. The user interface of this scene is shown in Fig. 4(b).
- **Argue scene:** Here, two users can engage in a process of *bilateral argumentation*, wherein they can propose aspects of the image which they like or dislike, in the form of tags. The aim is to convince the other user to align their opinions with yours, in terms of tag sizes. For example, one user might like the 'black and white' aspect of an image, whereas the other user dislikes it; one user can then pass this tag to the other user to request that they resize it. The user interface of this scene is shown in Fig. 4(c).
- **Vote scene:** Here, the decision is made to add an image to the group collection or not by *voting*. The user interface of this scene is shown in Fig. 4(d).

In the following section, the decision making criteria used in the WeCurate workflow are described.

#### 4. Collective decision making models

In a multiuser museum interactive system, it is not only important to model users and user preferences but also to assist them in making decisions. For example, the system could decide which artefact is worthy to be added to a group collection by merging user preferences (Yee-King et al., 2012); or it could decide whether the artefact is collectively accepted by a group of users by considering user evaluations about certain criteria of the artefact itself like in multiple criteria decision making (Ribeiro, 1996); or assist users in reaching agreements by argument exchange like in argument-based negotiation (Amgoud et al., 2012). These cases, that are essentially decision making problems, can be solved by defining different decision principles that take the preferences of the users into account and compute the decision of the group as a whole.

In the WeCurate system, agents base their decisions on two different models: preference aggregation and multiple-criteria decision making. The former is used to understand whether the users consider an image as interesting or not. To this end, each user expresses an *image preference* and a collective decision is made by aggregating the image preferences of all the users. The latter amounts to a collective decision made by discussion. Users exchange *image arguments* according to an *argument-based multiple criteria decision making* protocol.

UserAssistant agents assist the system and the users with several decisions and with an automatic updating mechanism in the different scenes. Namely:

- **Select scene:**
  - **Image interestingness:** Given the image preferences of all the users running in a select scene, the UserAssistant agent is responsible to decide whether the image (which is currently browsed) is interesting enough to be further discussed in a forum scene;



- **Forum scene:**
  - *Automatic image preference slider updater:* The UserAssistant agent updates the image preference slider of its user when the user rates the image by specifying a certain tag;
  - *Argue Candidate Recommender:* When a user decides to argue with another user, the UserAssistant agent recommends its user a list of possible candidates ordered according to the distance between their image preferences;
  - *Multi-criteria decision:* Given the image tags of all the users running in a forum scene, the UserAssistant agent is responsible to decide whether the image can be automatically added (or not) to the image collection without a vote being necessary;
- **Argue scene:**
  - *Automatic image preference slider updater:* The UserAssistant agent updates the image preference slider of its user when the user accepts an image tag proposed by the other user during the arguing;
  - *Argue agreement:* The UserAssistant agent ends the arguing among two users as soon as it detects that their image preferences are close enough.
- **Vote scene:**
  - *Vote counting:* The UserAssistant agent counts the votes expressed by the users running in a vote scene in order to decide if the image will be added (or not) to the image collection being curated.

For each scene, we describe the decision models into details.

#### 4.1. Select scene

The main goal of each user running in a select scene is to express a preference about the image currently browsed. When the scene ends, the UserAssistant agents compute an evaluation of the image, the *image interestingness* of the group of users by aggregating user preferences. The result of the aggregation is used to decide whether the users can proceed in a forum scene or whether a new select scene with a different image has to be instantiated.

##### 4.1.1. Preference aggregation

To formalise the decision making model based on preference aggregation, we introduce the following notation. Let  $\mathcal{I} = \{im_1, \dots, im_n\}$  be a set of available images where each  $im_j \in \mathcal{I}$  is the identifier of an image. The *image preference* of a user w.r.t. an image is a value that belongs to a finite bipolar scale  $S = \{-1, -0.9, \dots, 0.9, 1\}$  where  $-1$  and  $+1$  stand for 'reject' and 'accept' respectively. Given a group of  $n$  users  $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ , we denote the image preference of a user  $u_i$  w.r.t. an image  $im_j$  by  $r_i(im_j) = v_i$  with  $v_i \in S$ .

A *preference aggregator operator* is a mapping  $f_{agg} : S^n \rightarrow S$ , and  $f_{agg}$  is used to merge the preferences of a group of  $n$  users w.r.t. an image  $im_j$ . A generic decision criterion for making a decision about the interestingness of an image  $im_j$  can be defined as

$$\text{int}(im_j) = \begin{cases} 1 & \text{if } 0 < f_{agg}(\vec{r}) \leq 1 \\ 0 & \text{if } -1 \leq f_{agg}(\vec{r}) \leq 0 \end{cases} \quad (1)$$

where  $\vec{r} = \{r_1(im_j), \dots, r_n(im_j)\}$  is a vector consisting of the image preferences of  $n$  users w.r.t. an image  $im_j$ . Eq. (1) is a generic aggregator operator that can be instantiated using different functions for aggregating user preferences. In WeCurate, we have used three different preference aggregators that we describe as follows.

*Image interestingness based on arithmetic mean:* The *image interestingness* of a group of  $n$  users w.r.t. an image  $im_j$  based on arithmetic mean, denoted by  $f(\vec{r})$ , is defined as

$$f(\vec{r}) = \frac{\sum_{1 \leq i \leq n} r_i}{n} \quad (2)$$

Then, a decision criterion for the interestingness of an image  $im_j$ , denoted as  $\text{int}(im_j)$ , can be defined by setting  $f_{agg}(\vec{r}) = f(\vec{r})$  in Eq. (1). According to this definition, the system proceeds with a forum scene when  $\text{int}(im_j) = 1$ , while the system goes back to a select scene when  $\text{int}(im_j) = 0$ .

*Image interestingness based on weighted mean:* Each UserAssistant agent also stores the zoom activity of its user. The zoom activity is a measure of the user interest in a given image and, as such, it should be taken into account in the calculation of the image interestingness.

Let us denote the number of *image zooms* of user  $u_i$  w.r.t. an image  $im_j$  as  $z_i(im_j)$ . Then, we can define the total number of zooms for an image  $im_j$  as  $z(im_j) = \sum_{1 \leq i \leq n} z_i(im_j)$ . Based on  $z(im_j)$  and the  $z_i$ 's associated with each user, we can define a weight for the image preference  $r_i$  of user  $u_i$  as  $w_i = \frac{z_i}{z(im_j)}$ .

The *image interestingness* of  $n$  users w.r.t. an image  $im_j$  based on the weighted mean, denoted by  $f_{wm}(\vec{r})$ , can be defined as

$$f_{wm}(\vec{r}) = \frac{\sum_{1 \leq i \leq n} r_i w_i}{\sum_{1 \leq i \leq n} w_i} \quad (3)$$

Then, a decision criterion for the interestingness of an image  $im_j$  based on weighted mean, denoted as  $\text{int}_{wm}(im_j)$ , can be defined by setting  $f_{agg}(\vec{r}) = f_{wm}(\vec{r})$  in Eq. (1). The system proceeds with a forum scene when  $\text{int}_{wm}(im_j) = 1$ , while the system goes back to a select scene when  $\text{int}_{wm}(im_j) = 0$ .

*Image interestingness based on WOWA operator:* An alternative criterion for deciding whether an image is interesting or not can be defined by using a richer average operator such the Weighted Ordered Weighted Average (WOWA) operator (Torra, 1997).

The WOWA operator is an aggregation operator which allows us to combine some values according to two types of weights: (i) a weight referring to the importance of a value itself (as in the weighted mean), and (ii) an ordering weight referring to the values' order. Indeed, WOWA generalizes both the weighted average and the ordered weighted average (Yager, 1988). Formally, WOWA is defined as (Torra, 1997)

$$f_{wowa}(r_1, \dots, r_n) = \sum_{1 \leq i \leq n} \omega_i r_{\sigma(i)} \quad (4)$$

where  $\sigma(i)$  is a permutation of  $\{1, \dots, n\}$  such that  $r_{\sigma(i-1)} \geq r_{\sigma(i)} \forall i = 2, \dots, n$ ,  $\omega_i$  is calculated by means of an increasing monotone function  $w^*(\sum_{i \leq i} p_{\sigma(i)}) - w^*(\sum_{j < i} p_{\sigma(j)})$ , and  $p_i, w_i \in [0, 1]$  are the weights and the ordering weights associated with the values respectively (with the constraints  $\sum_{1 \leq i \leq n} p_i = 1$  and  $\sum_{1 \leq i \leq n} w_i = 1$ ).

We use the WOWA operator for deciding whether an image is interesting in the following way. Let us take the weight  $p_i$  for the image preference  $r_i$  of user  $u_i$  as the percentage of zooms made by the user (like above). As far as the ordering weights are concerned, we can decide to give more importance to image preference's values closer to extreme value such as  $-1$  and  $+1$ , since it is likely that such values can trigger more discussions among the users rather than image preference' values which are close to 0. Let us denote the sum of the values in  $S^+ = [0, 0.1, \dots, 0.9, 1]$  as  $s$ . Then, for each image preference  $r_i(im_j) = v_i$  we can define an ordering weight as  $w_i = \frac{r_i(im_j)}{s}$ . Please notice that the  $p_i$ 's and  $w_i$ 's defined satisfy the constraints  $\sum_{1 \leq i \leq n} p_i = 1$  and  $\sum_{1 \leq i \leq n} w_i = 1$ .

Then, a decision criterion for the interestingness of an image  $im_j$  based on WOWA, denoted as  $\text{int}_{\text{wowa}}(im_j)$ , can be defined by setting  $f_{\text{agg}}(\vec{r}) = f_{\text{wowa}}(\vec{r})$  in Eq. (1).

#### 4.2. Forum scene

The main goal of the users in a forum scene is to discuss an image, which has been considered interesting enough in a select scene, by pointing out what they like or dislike of the image through *image arguments* based on tags. During the tagging, the *overall image preference per user* is automatically updated. Whilst tagging is the main activity of this scene, a user can also choose to argue with another user in order to persuade him to adopt his own view (i.e. to “keep” or to “discard” the image). In such a case, a list of *recommended argue candidates* is retrieved. Finally, when a user is tired of tagging, he can propose the other users to move to a vote scene. In this case, an automatic *multi-criteria decision* is taken in order to decide whether the current image can be added or not to the image collection without a vote being necessary.

##### 4.2.1. Argument-based multiple criteria decision making

In our system each image is described with a finite set of *tags* or *features*. Tags usually are a convenient way to describe folksonomies (Hellin-Hobbs, 2010; Cosley et al., 2008; Cosley et al., 2009). In what follows, we show how weighted tags, that is, tags associated with a value belonging to a bipolar scale, can be used to define arguments in favor or against a given image and to specify a multiple criteria decision making protocol to let a group of users to decide whether to accept or not an image.

##### 4.2.2. Arguments

The notion of argument is at the heart of several models developed for reasoning about defeasible information (e.g. Dung, 1995; Pollock, 1992), decision making (e.g. Amgoud and Prade, 2009; Bonet and Geffner, 1996), practical reasoning (e.g. Atkinson et al., 2004), and modeling different types of dialogues (e.g. Amgoud et al., 2007; Prakken, 2005). An argument is a reason for believing a statement, choosing an option, or doing an action. In most existing works on argumentation, an argument is either considered as an abstract entity whose origin and structure are not defined, or it is a logical proof for a statement where the proof is built from a knowledge base.

In our application, *image arguments* are reasons for accepting or rejecting a given image. They are built by users when rating the different tags associated with an image. The set  $\mathcal{T} = \{t_1, \dots, t_k\}$  contains all the available tags. We assume the availability of a function  $\mathcal{F} : \mathcal{I} \rightarrow 2^{\mathcal{T}}$  that returns the tags associated with a given image. Note that the same tag may be associated with more than one image. A tag which is evaluated positively creates an *argument pro* the image whereas a tag which is rated negatively induces an *argument con* against the image. Image arguments are also associated with a weight which denotes the *strength* of an argument. We assume that the weight  $w$  of an image argument belongs to the finite set  $\mathcal{W} = \{0, 0.1, \dots, 0.9, 1\}$ . The tuple  $\langle \mathcal{I}, \mathcal{T}, \mathcal{S}, \mathcal{W} \rangle$  will be called a *theory*.

**Definition 4.1** (Argument). Let  $\langle \mathcal{I}, \mathcal{T}, \mathcal{S}, \mathcal{W} \rangle$  be a theory and  $im \in \mathcal{I}$ .

- An *argument pro*  $im$  is a pair  $((t, v), w, im)$  where  $t \in \mathcal{T}$ ,  $v \in \mathcal{S}$  and  $v > 0$ .
- An *argument con*  $im$  is a pair  $((t, v), w, im)$  where  $t \in \mathcal{T}$ ,  $v \in \mathcal{S}$  and  $v < 0$ .

The pair  $(t, v)$  is the *support* of the argument,  $w$  is its *strength* and  $im$  is its *conclusion*. The functions `Tag`, `Val`, `Str` and `Conc` return

respectively the tag  $t$  of an argument  $((t, v), w, im)$ , its value  $v$ , its weight  $w$ , and the conclusion  $im$ .

It is well-known that the construction of arguments in systems for defeasible reasoning is monotonic (see Amgoud and Besnard, 2009 for a formal result). Indeed, an argument cannot be removed when the knowledge base from which the arguments are built is extended by new information. This is not the case in our application. When a user revises his opinion about a given tag, the initial argument is removed and replaced by a new one. For instance, if a user assigns the value 0.5 to a tag  $t$  which is associated with an image  $im$ , then he decreases the value to 0.3, the argument  $((t, 0.5), w, im)$  is no longer considered as an argument and is completely removed from the set of arguments of the user and is replaced by the argument  $((t, 0.3), w, im)$ . To say it differently, the set of arguments of a user contains only one argument per tag for a given image.

In a forum scene, users propose, revise, and reject arguments about images by adding, editing and deleting bubble tags. Proposing a new argument about an image, for instance “I like the blue color very much”, is done by adding a new bubble tag “blue color” and increasing its size. When an argument of such a kind is created, is sent to all the users (taking part in the forum scene) and it is displayed in their screens as a bubble tag. At this point, the content of the image argument, e.g. the “blue color” tag, is implicitly accepted by the other users unless the corresponding bubble tag is deleted. However, the implicit acceptance of the argument does not imply that the value of the argument is accepted, which is assumed to be 0. This is because we assume that if someone sees a new tag and does not “act” on it, it means that she/he is indifferent w.r.t. that tag. The value of an argument is changed only when a user makes the bubble corresponding to the argument, bigger and smaller. On the other hand, the acceptance of arguments in an argue scene is handled in a different way as we shall explain in Section 4.3.

Since users will collectively decide by exchanging argument whether to accept or not an image, a way for analysing the opinions of the users w.r.t. the image is worthy to be explored.

##### 4.2.3. Opinion analysis

Opinion analysis is gaining increasing interest in linguistics (see e.g. Albert et al., 2011; Krishna-Bal and Saint-Dizier, 2010) and more recently in AI (e.g. Osman et al., 2006; Subrahmanian, 2009). This is due to the importance of having efficient tools that provide a synthetic view on a given subject. For instance, politicians may find it useful to analyse the popularity of new proposals or the overall public reaction to certain events. Companies are definitely interested in consumer attitudes towards a product and the reasons and motivations of these attitudes. In our application, it may be important for each user to know the opinion of a user about a certain image. This may lead the user to revise his own opinion.

The problem of opinion analysis consists of aggregating the opinions of several agents/users about a particular subject, called *target*. An opinion is a global rating that is assigned to the target, and the evaluation of some features associated with the target. Therefore, this amounts to aggregate arguments which have the structure given in Definition 4.1.

In our application, the target is an image and the features are the associated tags. We are mainly interested in two things. To have a synthetic view of the opinion of a given user w.r.t. an image and to calculate whether the image can be regarded as worthy to be accepted or not. In the first case, we aggregate the image arguments of a user  $u_i$  to obtain his overall image preference  $r_i^*$ . Instead, for deciding whether an image is accepted or rejected by the whole group we define a multiple criteria operator.

**Definition 4.2** (*Opinion aggregation*). Let  $\mathcal{U} = \{u_1, \dots, u_n\}$  be a set of users,  $im \in \mathcal{I}$  where  $\mathcal{F}(im) = \{t_1, \dots, t_m\}$ . The next table summarizes the opinions of  $n$  users.

Users/Tags	$t_1$	...	$t_j$	...	$t_m$	$im$
$u_1$	$(v_{1,1}, w_{1,1})$	...	$(v_{1,j}, w_{1,j})$	...	$(v_{1,m}, w_{1,m})$	$r_1^*$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_i$	$(v_{i,1}, w_{i,1})$	...	$(v_{i,j}, w_{i,j})$	...	$(v_{i,m}, w_{i,m})$	$r_i^*$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_n$	$(v_{n,1}, w_{n,1})$	...	$(v_{n,j}, w_{n,j})$	...	$(v_{n,m}, w_{n,m})$	$r_n^*$

The *aggregate* or *overall image preference* of a user  $u_i$  denoted by  $r_i^*(im)$  is defined as

$$r_i^*(im) = \frac{\sum_{1 \leq j \leq m} v_{ij} w_{ij}}{\sum_{1 \leq j \leq m} w_{ij}} \quad (5)$$

The multiple criteria decision operator can then be defined as

$$MCD(im) = \begin{cases} 1 & \text{if } \forall u_i, 0 \leq r_i^*(im) \leq 1 \\ -1 & \text{if } \forall u_i, -1 \leq r_i^*(im) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Note that the  $MCD$  aggregation operator allows three values: 1 (for acceptance),  $-1$  (for rejection) and 0 (for undecided). Therefore, an image  $im$  is automatically added to the image collection if it has been unanimously accepted by the users. On the contrary, the image is discarded if it has been unanimously rejected. Finally, if  $MCD(im) = 0$ , then the system is unable to decide and the final decision is taken by the users in a vote scene.

Notice that our definition of  $MCD$  captures the idea that a vote is needed only when users do not reach a consensus in the forum and argue scenes.<sup>2</sup>

#### 4.2.4. Overall image preference per user

When a user rates an image  $im$  by specifying a new tag or by updating a tag already specified, his overall image preference is automatically updated by computing  $r_i^*(im)$ .

#### 4.2.5. Argue candidate recommender

In order to recommend an ordered list of argue candidates to a user willing to argue, the distance between the overall image preferences per user (Eq. (5)) can be taken into account.

Let  $u_i$  be a user willing to argue and  $r_i^*(im)$  be his overall image preference. Then, for each  $u_j$  (such that  $j \neq i$ ) we can define the *image preference distance* of user  $u_j$  w.r.t. user  $u_i$ , denoted by  $\delta_{ji}(im)$ , as

$$\delta_{ji}(im) = \{ \text{abs}(r_j^*(im) - r_i^*(im)) \mid (r_j^*(im) < 0 \wedge r_i^*(im) \geq 0) \vee (r_j^*(im) \geq 0 \wedge r_i^*(im) < 0) \} \quad (7)$$

Then, an argue candidate for user  $u_i$  for an image  $im$  is

$cand_i(im) = \{u_j \mid \max \{\delta_{ji}(im)\}\}$ . The ordered list of argue candidates can be defined by ordering the different  $\delta_{ji}(im)$ .

#### 4.3. Argue scene

The main goal of two users running in an argue scene is to try to reach an agreement on keeping or discarding an image by exchanging image arguments. The argue scene defines a bilateral argumentation protocol. The formal protocol is presented at the end of the section and it works as follows:

- the two users tag the image by means of image's tags (like in the forum scene), but, they can also propose image tags to the other user:
  - while tagging, their overall image preferences are automatically updated;
- a user proposes an image tag to the other user who can either accept or reject it:
  - if the user accepts the image tag proposed, then their overall image preferences are automatically updated:
    - \* if an *argue agreement* is reached, then the argue scene stops;
    - \* otherwise, the argue scene keeps on;
  - if the user rejects the image tag proposed, then the argue scene keeps on.

Both users can also decide to leave the argue scene spontaneously.

Whilst in a forum scene, an argument is implicitly accepted unless the corresponding bubble tag is deleted, in the above protocol, when a user proposes an argument to another user, the second user can accept or reject that argument by clicking on the bubble tag representing the argument and selecting an accept/reject option. The user who accepts the argument accepts not only the content of the argument but also its value. Previous arguments over the same tag (if they exist) are overwritten.

The different way in which an argument is accepted or rejected in a forum and an argue scene, is motivated by the different, although related, intended goals of the two scenes. Whilst the goal of the forum scene is to develop a sense of community discourse around an image (and the deletion a bubble tag of another user can foster the creation of new arguments), the goal of the argue scene is to support a “private” bilateral negotiation protocol that lets a user to persuade another one about the specifics of an image.

##### 4.3.1. Overall image preference per user

The overall image preference of a user in an argue scene is automatically updated by computing  $r^*(im)$  (see subsection 4.2.4).

##### 4.3.2. Argue agreement

Informally, an *argue agreement* is reached when the image preferences of the two users agree towards “keep” or “discard”. Let  $r_i^*(im)$  and  $r_j^*(im)$  be the image's preferences of user  $u_i$  and  $u_j$  respectively. Then, a decision criterion for deciding whether an argue agreement is reached can be defined as:

$$\text{argue}(im) = \begin{cases} 1 & \text{if } (0 \leq r_i^*(im) \leq 1 \wedge 0 \leq r_j^*(im) \leq 1) \vee \\ & (-1 \leq r_i^*(im) < 0 \wedge -1 \leq r_j^*(im) < 0) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Therefore, an argue scene stops when  $\text{argue}(im) = 1$ . Instead, while  $\text{argue}(im) = 0$ , the argue scene keeps on until either  $\text{argue}(im) = 1$  or the two users decide to stop arguing. The “otherwise” case covers the situation in which the overall image preferences of two users are neither both positive nor negative. This corresponds to a disagreement situation and to the case in which the users should keep arguing. Therefore, the system should

<sup>2</sup> Although it is quite probable that if users are heterogeneous the obtained value of  $MCD$  will be 0, during our trials at the Horninam museum, most of the people using WeCurate were groups of friends and families. This lowered the probability that their views diverged, and we wanted to have a decision making model that let them vote only in case they were not unanimously agreeing on what to do. Please notice that, since the  $MCD$  is a decision criterion run by the agents participating to the EI, we can obtain a different behaviour of the group by plugging in another decision model.



not interrupt the argue protocol which can be stopped by one of the users as mentioned in Section 4.3.

The reader might notice that user image preferences with a value of 0 and  $-0.1$ , although mathematically very close, contribute to make different decisions. This view is justified by the fact that we categorise the satisfaction and dissatisfaction of a user w.r.t an image taking a possibility theory approach to user preference representation and fusion into account (Benferhat et al., 2006). According to this approach, user preferences are modeled in terms of a finite bipolar scale in which values in the range  $[1, 0.9, \dots, 0.1, 0]$  represent a set of satisfactory states (with 1 being a state of full satisfaction and 0 a state of indifference), while values in the range  $(0, -0.1, \dots, -0.9, -1]$  capture states of dissatisfaction (with  $-0.1$  being a state of low dissatisfaction and  $-1$  being a state of maximum dissatisfaction). Therefore, according to this categorisation,  $-0.1$  is a state of dissatisfaction, while 0 is not. This is why  $-0.1$  and 0 are accounted as a negative and a positive value in the definition of *argue* respectively.

#### 4.4. Vote scene

The main goal of the users running in a vote scene is to decide by vote to add or not an image to the image collection. This decision step occurs when the automatic decision process at the end of the forum scene is unable to make a decision.

In a vote scene, each user vote can be “yes”, “no”, or “abstain” (in case that no vote is provided). Let  $v_i \in \{+1, 0, -1\}$  be the vote of user  $u_i$  where  $+1$  = “yes”,  $-1$  = “no”, and  $0$  = “abstain” and let  $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$  be the set of votes of the users in a vote scene. Then, a decision criterion for adding an image or not based on *vote counting* can be defined as

$$\text{vote}(im_j) = \begin{cases} 1 & \text{if } \sum_{1 \leq i \leq n} v_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Therefore, an image  $im_j$  is added to the image collection if the number of “yes” is greater or equals than the number of “no”. In the above criterion, a neutral situation is considered as a positive vote.<sup>3</sup>

#### 4.5. Agent interaction protocol

In the previous sections, we have mainly presented the architecture of the system and the reasoning part of the agents in the system. In what follows we provide the interaction protocol followed by the agents in the different scenes. We describe the negotiation protocol that allows agents to make joint decisions. The idea is the following. Whenever a sufficient number of UserAssistant agents have logged in the system, the EIManager starts a select scene. Each user will zoom into an image and express an image preference. When a user decides to go to a forum scene, its UserAssistant agent computes the group preference by means of a preference aggregator. Based on this result ( $\text{int}(im)$ ) the EIManager decides whether to go to a forum or to go back to a select scene (with a different image). In the forum scene, each user will express his opinion about the image by specifying image arguments (as in Definition 4.1) via the system interface (see Section 5). Agents provide to their respective users a report on the aggregated opinion of the other users. Users may consider this information for revising their own opinions. In case all agents agree, that is,  $\text{MCD}(im) = 1$  (reps. disagree, that is,  $\text{MCD}(im) = -1$ ) on the overall rating of the image, then the image is added (resp. not added) to a group collection and another instance of a select scene is started. During the discussion, pairs of users may engage in private dialogues where they exchange arguments about the image. The exchanged

arguments may be either the ones that are built by the user when introducing his opinion or new ones. A user may add new tags for an image. When the disagreement persists ( $\text{MCD}(im) = 0$ ), the users will decide by voting.

In what follows,  $\mathcal{U} = \{u_1, \dots, u_n\}$  is a set of users, and  $\text{Args}^t(u_i)$  is the set of arguments of user  $u_i$  at step  $t$ . At the beginning of a session, the sets of arguments of all users are assumed to be empty (i.e.,  $\text{Args}^0(u_i) = \emptyset$ ). Moreover, the set of images contains all the available images in the database of the museum, that is  $\mathcal{I}^0 = \mathcal{I}$ . We assume also that a user  $u_i$  is interested in having a joint experience with other users. The protocol uses a communication language based on four locutions:

- **Invite**: it is used by a user to invite a set of users for engaging in a dialogue.
- **Send** is used by agents for sending information to other agents.
- **Accept** is used mainly by users for accepting requests made to them by other users.
- **Reject** is used by users for rejecting requests made to them by other users.

##### Interaction protocol:

1. **Send**(EIManager,  $\mathcal{U}$ , **SelectScene**) (the EIManager starts a select scene).
2. **Send**(MediaAgent,  $\mathcal{U}$ ,  $\text{Rand}(\mathcal{I}^t)$ ) (the Media Agent select an image from the museum database and sends it to all the UserAssistant agents).
3. Each UserAssistant agent displays the image  $\text{Rand}(\mathcal{I}^t)$  and each user  $u_j \in \mathcal{U}$ :
  - (a) Expresses an image preference  $r_j(\text{Rand}(\mathcal{I}^t)) \in \mathcal{S}$ .
  - (b) When a user  $u_j$  is sure about his preference, he clicks on the “Go To Discuss” button in the WeCurate interface.
  - (c) **Send**(UserAssistant <sub>$j$</sub> , EIManager,  $f_{\text{agg}}(\vec{r})$ ) (the UserAssistant agent of  $u_j$  computes the group preference  $f_{\text{agg}}(\vec{r})$  and sends it to the EIManager).
4. If ( $\text{int}(\text{Rand}(\mathcal{I}^t)) = 0$ ), then  $\mathcal{I}^{t+1} = \mathcal{I}^t \setminus \{\text{Rand}(\mathcal{I}^t)\}$  and go to Step 1.
5. If ( $\text{int}(\text{Rand}(\mathcal{I}^t)) = 1$ ), then **Send**(EIManager,  $\mathcal{U}$ , **ForumScene**) (the EIManager starts a forum scene).
6. Each UserAssistant agent displays the image  $\text{Rand}(\mathcal{I}^t)$  and its tags (i.e.,  $t_i \in \mathcal{F}(\text{Rand}(\mathcal{I}^t))$ ). [Steps 7 and 8 can happen in parallel]
7. Each user  $u_j \in \mathcal{U}$ :
  - (a) creates image arguments. Let  $\text{Args}_j^t = \text{Args}_j^{t-1} \cup \{((t_i, v_i), w_i), \text{Rand}(\mathcal{I}^t)) \mid t_i \in \mathcal{F}(\text{Rand}(\mathcal{I}^t))\}$  be the set of arguments of user  $u_j$  at step  $t$ .
  - (b) The UserAssistant agent of  $u_j$  computes his overall image preference and the one of the other users  $r_i^*(\text{Rand}(\mathcal{I}^t))$ .
  - (c) The user  $u_j$  may change his opinion in light of  $r_i^*(\text{Rand}(\mathcal{I}^t))$ . The set  $\text{Args}_j^t$  is revised accordingly. All the arguments that are modified are replaced by the new ones. Let  $\mathcal{T}' \subseteq \mathcal{F}(\text{Rand}(\mathcal{I}^t))$  be the set of tags whose values are modified. Therefore,  $\text{Args}_j^t = (\text{Args}_j^{t-1} \setminus \{((t, v), w), (\text{Rand}(\mathcal{I}^t)) \in \text{Args}_j^{t-1} \mid t \in \mathcal{T}'\}) \cup \{((t, v'), w'), \text{Rand}(\mathcal{I}^t)\} \mid t \in \mathcal{T}'$ .  $r_i^*(\text{Rand}(\mathcal{I}^t))$  is calculated everytime the set image argument is modified.
  - (d) When the user  $u_j$  is sure about his preferences, he clicks on the “Go To Vote” button in the WeCurate interface.
  - (e) **Send**(UserAssistant <sub>$j$</sub> , EIManager,  $r_i^*(\text{Rand}(\mathcal{I}^t))$ ) (the UserAssistant agent sends  $r_i^*(\text{Rand}(\mathcal{I}^t))$  to the EIManager).
8. For all  $u_j, u_k \in \mathcal{U}$  such that  $\delta_{kj}(\text{Rand}(\mathcal{I}^t)) > 0$  then:
  - (a) **Invite**( $u_j, \{u_k\}$ ) (user  $u_j$  invites user  $u_k$  for a private dialogue).
  - (b) User  $u_k$  utters either **Accept**( $u_k$ ) or **Reject**( $u_k$ ).
  - (c) If **Accept**( $u_k$ ), then **Send**(EIManager,  $\{u_i, u_k\}$ , **ArgueScene**).

<sup>3</sup> This assumption is made to avoid an *undecided* outcome at this decision step.

- (d)  $\text{Send}(u_j, \{u_k\}, a)$  where  $a$  is an argument,  $\text{Conc}(a) = \text{Rand}(\mathcal{I}^t)$  and either  $a \in \text{Args}_j^t$  or  $\text{Tag}(a) \notin \mathcal{I}$  (i.e., the user introduces a new argument using a new tag).
- (e) User  $u_k$  may revise his opinion about  $\text{Tag}(a)$ . Thus,  $\text{Args}_k^t = (\text{Args}_k^t \setminus \{((\text{Tag}(a), v), \text{Rand}(\mathcal{I}^t))\}) \cup \{((\text{Tag}(a), v'), \text{Rand}(\mathcal{I}^t)) \mid v' \neq v\}$ .
- (f) If  $(\text{argue}(\text{Rand}(\mathcal{I}^t)) = 0 \wedge \text{not exit})$ , then go to Step 8 (d) with the roles of the agents reversed.
- (g) If  $(\text{argue}(\text{Rand}(\mathcal{I}^t)) = 1 \vee \text{exit})$ , then go to Step 7.
9. If  $(\text{MCD}(\text{Rand}(\mathcal{I}^t)) = -1)$ , then  $\mathcal{I}^{t+1} = \mathcal{I}^t \setminus \{\text{Rand}(\mathcal{I}^t)\}$  and go to Step 1.
10. If  $(\text{MCD}(\text{Rand}(\mathcal{I}^t)) = 1)$ , then  $\text{Rand}(\mathcal{I}^t)$  is added to the group collection,  $\mathcal{I}^{t+1} = \mathcal{I}^t \cup \{\text{Rand}(\mathcal{I}^t)\}$  and go to Step 1.
11. If  $(\text{MCD}(\text{Rand}(\mathcal{I}^t)) = 0)$ , then  $\text{Send}(\text{ElManager}, \mathcal{U}, \text{VoteScene})$  (the ElManager starts a vote scene).
12. Each user  $u_j \in \mathcal{U}$ :
  - (a) expresses a vote  $v_j(\text{Rand}(\mathcal{I}^t))$ .
  - (b)  $\text{Send}(\text{UserAssistant}_j, \text{ElManager}, v_j(\text{Rand}(\mathcal{I}^t)))$ .
13. If  $(\text{vote}(\text{Rand}(\mathcal{I}^t)) = 1)$ , then  $\text{Rand}(\mathcal{I}^t)$  is added to the group collection,  $\mathcal{I}^{t+1} = \mathcal{I}^t \cup \{\text{Rand}(\mathcal{I}^t)\}$  and go to Step 1.
14. If  $(\text{vote}(\text{Rand}(\mathcal{I}^t)) = 0)$ , then  $\mathcal{I}^{t+1} = \mathcal{I}^t \setminus \{\text{Rand}(\mathcal{I}^t)\}$  and go to Step 1.

It is worth mentioning that when a user does not express opinion about a given tag, then he is assumed to be indifferent w.r.t. that tag. Consequently, the value 0 is assigned to the tag.

Note also that the step 8 is not mandatory. Indeed, the invitation to a bilateral argumentation is initiated by users who really want to persuade their friends.

The previous protocol generates dialogues that terminate either when all the images in the database of the museum are displayed or when users exit. The outcome of each iteration of the protocol may be either an image on which all users agree or disagree to be added to the group collection.

## 5. User interface

The user interface provides a distinct screen for each scene, as illustrated in Figs. 4(a–d). It communicates with the UserAssistant agent by sending a variety of user triggered events which are different in each scene. The available user actions in each scene are shown in Fig. 1. The state of the interface is completely controlled by the UserAssistant agents, which send scene snapshots to the interface whenever necessary, e.g. when a new tag is created. Some low level details of the method of data exchange between interface and UserAssistant agents are provided in the next section.

The interface is the second iteration of a shared image browsing interface, designed to include desirable features highlighted by a user trial of the first iteration (see Hazelden et al., 2012 for more details). Desirable features include standard usability such as reliability, speed and efficiency, awareness of the social presence of other users and awareness of the underlying workflow. Given the social nature of the system, social presence, where users are aware of each others' presence and actions as well as a shared purpose and shared synchronicity is of especial interest.

## 6. Adopted technologies

The p2p EI is implemented on top of *FreePastry*, a free and open-source library that implements peer-to-peer networks (Rowstron and Druschel, 2001), and AMELI, a general-purpose middleware (i.e. set of institutional agents) that enables the execution of the EI. Whilst *FreePastry* provides several useful

features such as the routing of messages, or the possibility to create broadcast messages, AMELI enables agents to act in an EI and controls their behaviour. The institutional agents composing AMELI load institution specifications as XML documents generated by ISLANDER (Esteve et al., 2002), a graphical editor for EI specifications. AMELI is composed of three layers: a communication layer, which enables agents to exchange messages, a layer composed of the external agents that participate in an EI, and in between a social layer, which controls the behaviour of the participating agents. The social layer is implemented as a multi-agent system whose institutional agents are responsible for guaranteeing the correct execution of an EI according to the specification of its rules. User Assistant agents are implemented as Java programs extending an existing Java agent that abstracts away all the underlying communication protocols. More details on the p2p EI implementation can be found in de Jonge et al., 2013.

The user interface is implemented using Javascript drawing to an HTML5 canvas element, which is a cross platform and plug-in free solution. The Interface does not communicate directly with the institutional agents since it is not a part of the *FreePastry* network. Instead, the interface sends events formatted as JSON to the User Assistant agent which hosts an HTTP server. The User Assistant agents pick up the event queue, then in turn generate scene snapshots in JSON format which are sent to the interface. Scene snapshots are used to define the state of the interface.

One advantage of this queued event and snapshot model with regard to evaluation is that all interface events and interface state snapshots are stored in the system for later inspection. This allows a complete, interactive reconstruction of activity of the users and the agents for qualitative analysis as well as providing a lot of data for quantitative analysis. In the next section we describe how this data was analysed.

## 7. Evaluation

The objective of the evaluation is twofold. First, to determine the interactions and the social awareness perceived by the social groups using our system.

Second, to test to what extent the decision models adopted by the agents were good predictors of user behaviour e.g. to decide whether users add an image to the group collection by analysing user preferences in a select scene or the arguments exchanged in a forum scene.

### 7.1. Method and data

The WeCurate system was set up as an interactive exhibit in a major London museum, supported by the research team.<sup>4</sup> The museum provided 150 images from their collection; the task for the social groups interacting with the system was to decide which of these images they would like to have as a postcard, via the curation process.

Multiple sources of qualitative and quantitative data were collected. Participants were filmed during the activities and their interactions with the system was recorded in timestamped logs. Data was gathered and cross referenced from adhoc observation of the trials themselves, inspection of the video footage, transcription of the interviews and the system log files.

The ages of participants ranged from 4 years (with assistance) to 45 years. The average time each group used the WeCurate system was 5 min 38 s, the longest session logged was 21 min 16 s.

<sup>4</sup> A video of the interactive exhibit and the description of the system is available at <http://www.youtube.com/watch?v=LzZ1EQS0-hQ>.

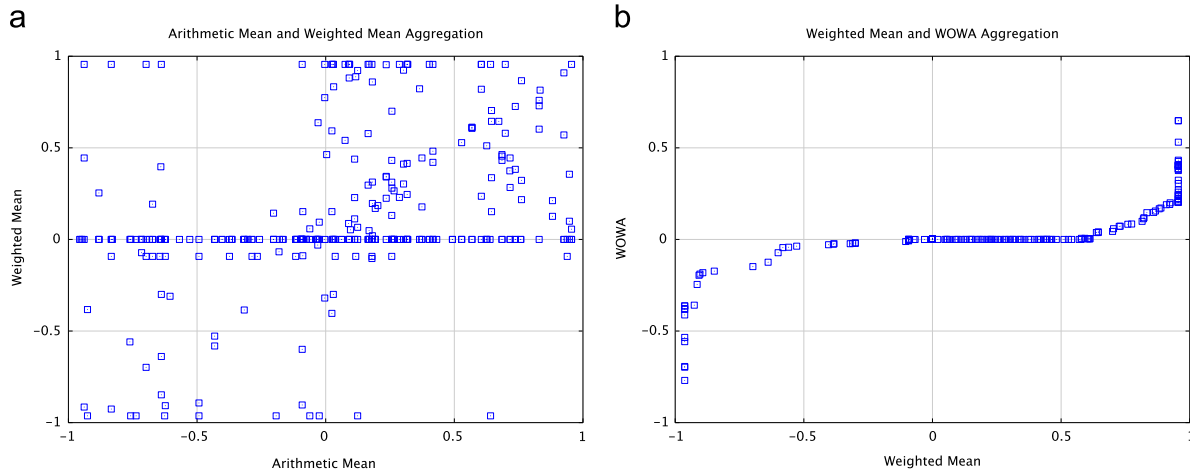


Fig. 5. The WeCurate operators in the select scene. (a) The Arithmetic and Weighted Mean Operators. (b) The Weighted Mean and WOWA Operators.

## 7.2. Community interaction analysis

For the social interaction analysis, the evaluation uses a Grounded Theory (GT) approach to code data from multiple sources to build an account of use (Glaser and Strauss, 1967; Glaser, 2008). GT enables a more speculative and emergent approach to rationalising the findings of the analysis. Of particular interest is the communication and discussion about the artefacts/images presented by the system, and whether the shared view supports an awareness of social action. The results of the community interaction analysis are presented in Hazelden et al. (2013) in a detailed way, here we only summarise the salient points:

- *Dynamic between adults and between parents and children:* Of the adult only sessions, 70% featured some degree of laughter and playful comments, these included reactions to another participant deleting a newly created tag, or commenting on the content of a tag. Consequently for the adults, the creation of a tag, or modifying a group member's tag was often perceived as a playful action. 60% of the adult's sessions also featured an attempt by at least one of the participants to synchronise their actions with the group (i.e. not clicking "Go To Discuss"/ "Go To Vote" until others were ready to move to the next image/vote). Aside from the positive communication among the adults, there were instances in 60% of these sessions where a participant expressed an opinion or asked for an opinion and no one responded. The lack of acknowledgement of group members comments could indicate that the participants were too engaged with the task and therefore did not register the comment, or they simply chose to ignore the group member. The social dynamic between parent and child was dominated by adult initiated action whereby 89% of the interactions related to the adult driving the child's comprehension. Of the adult initiated behaviour, 40% was directing the child's action and attention, and 45% was requesting an opinion about the image from the child.
- *Discussion of task, image and museum artefacts:* The questionnaire showed that 56% reported feeling as if they had a full discussion, while 23% reported that they did not (21% did not comment). Whilst it is encouraging that a majority believed they had a rich debate about the images in the system, as this a key aspect of the design and use, a more significant margin would be preferable. Of more concern is that in 30% of the sessions observed (with both adults and children) there was no discussion between the participants using separate devices, and in only one of these sessions did the children talk to each other

(in all other sessions they conversed solely with their parent). The absence of discussion could be partially accounted for by the parents preoccupation with supporting their child.

- *Social awareness via the system:* When reporting on their ability to express an opinion of the image in the questionnaire, 73% of participants felt they were able to express a preference in the select scene, and 81% reported that they could express opinions via the forum scene using the tags. This suggests that the participants felt they were able to communicate their preferences via the WeCurate Interface. The social group did appear to have some influence over individual's decision making, whereby 42% reported changing their decision as a consequence of seeing other's actions.

In what follows, we will focus on the analysis of the decision making models employed by the agents.

## 7.3. Agent decision models analysis

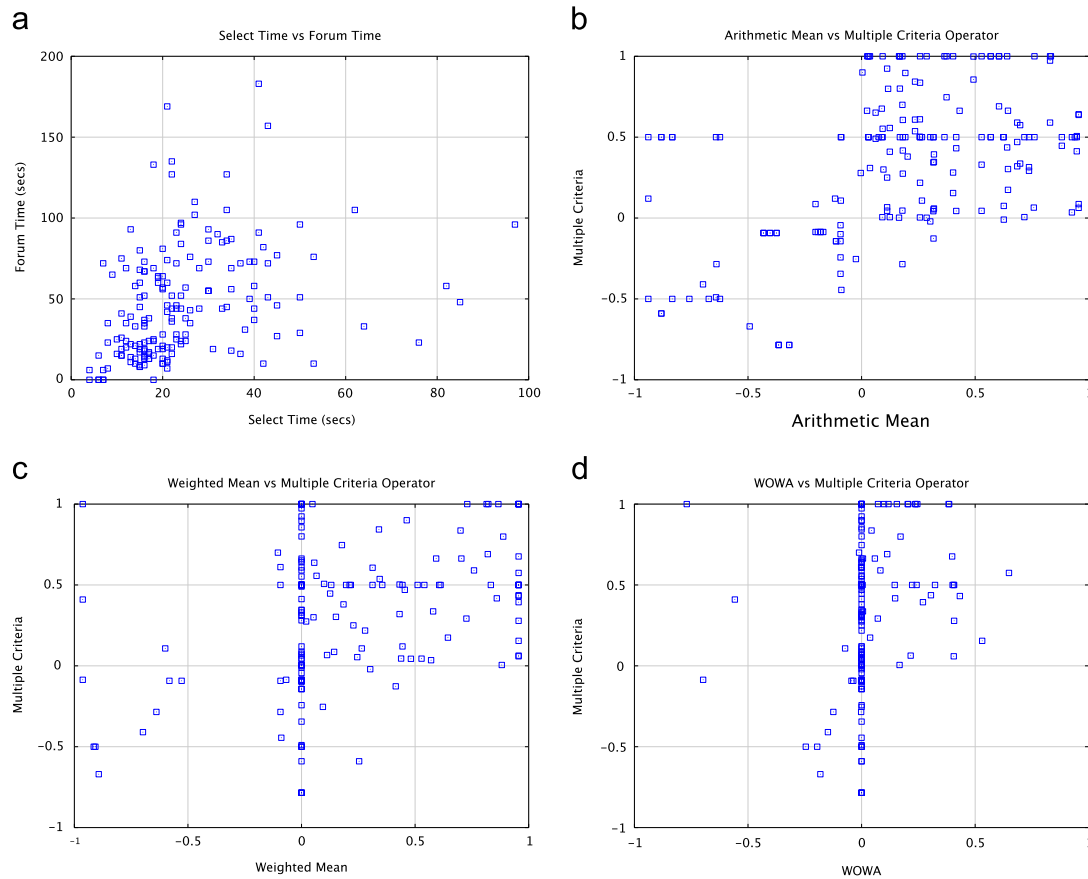
The observations provided a dataset for assessing the different types of agent decision models. To this end, we compare the decision criteria of the agents (Section 4) w.r.t. the final decision of the users in the vote scene.

The dataset analysed consists of 224 observations about image evaluations in the different WeCurate scenes. The images evaluated were selected from a finite set of 150 images, browsed during 165 sessions in which groups up to 4 users participated. 130 images were chosen from the original set and the 73.1% was finally added to the group collection. Each image was seen from 1 to 4 times during the different sessions. Among the 224 observations, 176 corresponded to image evaluations in which an image was added to the group collection and 48 in which an image was rejected by voting.

### 7.3.1. Select scene

For the analysis of the select scene, we considered the *number of users*, the *time spent* in the select scene, the *zoom activity* (number of zooms), the *image interestingness* computed on the basis of the three operators, and the different decision making criteria used by the agents ( $int$ ,  $int_{wm}$  and  $int_{wowa}$ ).

As general statistics, we observed that the shortest and longest select scene respectively took 7 and 105 s, with an average of 26 s for deciding to accept an image and 22 s for rejecting it. Therefore, it seems that the decision of disliking an image took slightly less than the decision of liking it. As far as the zoom activity is concerned, almost 50% of the select scenes did not have any zoom



**Fig. 6.** Relations between select and forum scene. (a) Select scene vs forum scene time. (b) The Arithmetic Mean and MCD. (c) The Weighted Mean and MCD. (d) The WOWA and the MCD.

activity. This could let us think that users did not zoom because they were not aware about this functionality. On the other hand, by looking at the cases with and without zoom activity, we appreciated that the lack of zoom activity corresponded to select scenes in which the image was finally rejected by the agents. Among those evaluations in which the zoom was used, the 74.4% classified the image as interesting, against the 55% in which the zoom activity was 0. Therefore, the zoom activity can be considered a positive measure of the users' activity w.r.t the image interestingness.

We also observed that there exists a significant positive correlation between different variables in the select scene.

First, a positive correlation related to the number of users versus the time spent in the select scene, the number of users versus the zoom activity, and the time spent in the select scene versus the zoom activity. These results can suggest us that users felt more engaged in using the application when other users were connected. This is in agreement with the kind of socio-technical system we implemented, where each user is aware of the activity of other users and social activities among users are stressed.

Second, the correlation of the zoom activity versus the image interestingness computed w.r.t. the different operators tell us that the algorithms used to compute the image interestingness were consistent w.r.t. the zoom activity of the users. Indeed, in the case of the arithmetic mean, the correlation with the zoom activity is not significant, while for the weighted mean and WOWA operators, which are zoom dependent, positive correlations, indicate that the number of zooms matters as expected.

Since not all the operators adopted were taking the zoom activity into account, it is interesting to compare them w.r.t the way they classify an image. Fig. 5 shows two graphics which

represent the relation between the arithmetic mean versus the weighted mean operator (Fig. 5(a)), and the weighted mean versus the WOWA operator (Fig. 5(b)).

In Fig. 5(a), it can be noticed that, although the values computed by the two operators correlate, the weighted mean operator classified as not interesting several images that the arithmetic mean considered interesting (because the zoom activity for those images was 0). Apart from those values, the two operators had a pretty good concordance since they classified most of the images in a similar way (see the top-right quadrant for class 1 and bottom-left quadrant for class 0), with the exception of some of them belonging to opposite (0 versus 1) classifications. This inconsistency can be explained by thinking about those cases in which small weights were associated to several positive user preferences and high weights were associated to few negative preferences, or vice-versa.

On the other hand, Fig. 5(b), reveals a very good concordance between the weighted mean and the WOWA operators since these operators classified images almost in the same way. This is somehow expected since both operators rely on the zoom activity. Nevertheless, the WOWA operator tends to flat low weighted mean values towards the 0 and to keep those values closer to extreme values +1 and -1.

### 7.3.2. Forum scene

For the analysis of the forum scene, we considered the *number of users*, the *time spent* in the forum scene, the *zoom activity*, the *tag activity* (the tags added, edited and deleted), the *comments*, the *forum preference*, and the multiple criteria operator MCD. By means of this operator, the agents classified an image as a good (1), a bad



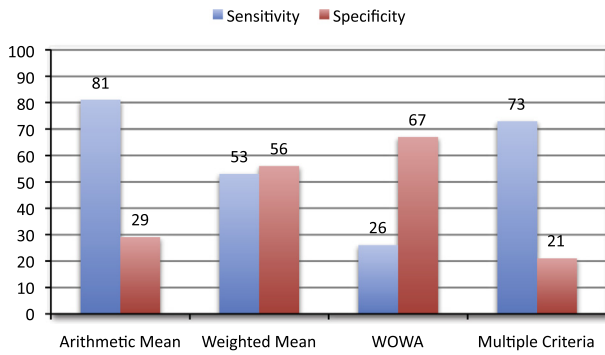


Fig. 7. Sensitivity and Specificity of the WeCurate operators used in the forum and select scenes w.r.t. the vote.

(−1), or a neutral (0) candidate to be added to the group collection.

We observed that the shortest and longest forum scene took 15 s and 210 s respectively, with an average of 55 s for those sessions (151) in which the agents recommended to add an image, 67 s for those sessions (21) in which the agents did not recommend it, and 15 s for those cases (52) in which the agents could not decide. Although we observed that the zoom and chat activities were quite low (6% of all the observations), it is interesting to notice that users were more engaged in the tag activity (85% of all the observations). In fact, a significant positive correlation exists between the time spent in the forum scene and all kinds of tag activities. Moreover, the activity of editing a tag is also positively correlated with the forum preference, since this value is computed on the basis of users' tags. Surprisingly, we discovered that the number of users and the time spent in the forum scene correlate in a weak way. This can be justified by thinking that many users already had a pretty clear idea of whether they liked or disliked the image and they tended to go to the vote scene without discussing it.

On the other hand, there exists a significant correlation between the number of users and the tag activity. This can be interpreted in two ways. First, we can expect that more users were likely to perform more tag activity. Second, it is also possible that users were more involved in tagging because they were aware of the tag activity of the other users (social awareness), and they felt more engaged.

Since each forum scene happened after a select scene in which an image was classified as interesting or not, it is worthy to look at the relation among the two scenes. First, we observed that the time spent in the select scene and the time spent in the forum scene are significantly correlated (Fig. 6). This was probably due to the fact that those images about which users were more undecided required more time to set a select and a forum preference. Second, we can draw a relation between the evaluations in the select and forum scene (Fig. 6(b–d)). Although the computation of the preference w.r.t. the images was based on different activities, that are, the aggregation of users' preferences in the select scene and the multiple criteria aggregation (MCD) of image tags' evaluations in the forum scene, it is interesting to observe how the different operators classified the images in a consistent way. Those values which are not in concordance correspond to those sessions of the forum scene in which users revised their opinions about an image chosen in the select scene. However, these cases represented a small percentage.

### 7.3.3. Vote scene

In the vote scene, users finally decided whether to add or not an image, browsed in the whole curation process, to the group collection. Therefore, it is interesting to compare this final decision

w.r.t. the decisions made by the agents in the forum and in the select scene.

To this end, we can measure the performance of our image classifiers in the forum and in the select scene in terms of sensitivity and specificity. In our case, the sensitivity of our operators refers to the capability of identifying good candidate images in the select and in the forum scene. On the other hand, the specificity is the capability of discriminating uninteresting images that finally were not voted.

For this analysis, we have considered the `vote` decision criterion (Eq. (9)), the `MCD` criterion (Eq. (6)) and, the decision criteria w.r.t. the image interestingness computed by `int`, `intwm`, and `intwowa`. Among the 224 image evaluations, 176 finally received a positive vote, while 48 a negative one. Fig. 7 shows the sensitivity and the specificity measures for the classifiers in the select and in the forum scenes.

As far as the classification in the select scene according to the three operators is concerned, we have observed the following. For the arithmetic mean, among those 176 observations that contained a positive vote, 34 of them were classified as false negatives (the image was accepted in the vote scene but not in the select scene), and 142 were classified as true positives in the select scene. Therefore its sensitivity is of 81%. Regarding its specificity, we have observed that in 48 observations, 34 of them were classified as false positive in the select scene (the image was chosen in the select but not in the vote); therefore its specificity is of 29%. On the other hand, the weighted mean and the WOWA show a sensitivity and a specificity of 53%, 56% and 26%, 67% respectively.

Within the forum scene, among those 176 observations that contained a positive vote, only 11 of them were classified as false negative, and 128 were classified as true positives in the forum scene. The remaining 37, would have required a vote anyway, since they remained unclassified in the forum scene. This gives us a sensitivity of 73%. Regarding the specificity, we have observed that in 48 observations, 23 of them were classified as false positives in the forum scene (the image was chosen in the forum but not in the vote), and 10 classified as true negatives. The remaining 15 would have required a vote anyway. This gives us a specificity of 21%.

## 8. Discussion

The analysis performed suggests that in the select scene, the arithmetic mean operator was not very sensitive at the moment of classifying the images, and for this reason, more images tended to go through the curation process, although they were finally rejected by voting. Instead, the weighted mean and the WOWA operators, since they depend on the number of zooms, and consequently, on the user activity, were more restrictive when selecting images. Indeed, they both are good classifiers with respect to the images that were finally voted. The WOWA operator, since it is more sensible to values closer to 0 (see Fig. 5b), discriminated too much in the selection of the images (26% of sensitivity). Thus, on the one hand, we can say that the weighted mean operator is a better image classifier than the arithmetic mean and the WOWA operators, which respectively are too weak and too strong with respect to the images they select. On the other hand, these results also suggest that a combination of the agents' decision models could enhance the user experience in using the system. For instance, by using the arithmetic mean operator to select images in the selection scene, but to finally vote only those images which are not discarded by the weighted mean or by the WOWA.

As far as the forum scene is concerned, the `MCD` operator categorised images that were finally voted in a pretty good way (73% of sensitivity). Its specificity, however, reveals that images

were classified as not worthy to be added to the group collection before the vote in few forum evaluations. Interpreting this result is difficult, but one possible explanation is that the vote scene was triggered hastily by those users who liked an image, preventing those who were changing their opinion during the discussion of the forum scene from expressing that change before moving to the voting scene.

## 9. Related work

Our work relates not only to research topics such as preference aggregation, argumentation and environments for multi-agent systems, but also to systems that allow realtime multiuser collaborations in cultural heritage and other domains.

As far as preference aggregation is concerned, our categorisation of positive and negative user preferences – to capture degrees of likeness and dislikeness of a user w.r.t. an image – is influenced by Benferhat et al. (2006), which proposes a bipolar fusion operator for merging user preferences in the possibilistic logic setting. According to this approach, the problem of deciding what is collectively accepted by a set of users can be handled by means of an aggregation function on the whole set of positive and negative preferences (represented in terms of possibility distributions) of a group of agents. On the other hand, one of our preference aggregation operators, used to decide whether an image is accepted or rejected in a select scene, is based on the Weighted Ordered Weighted Average (WOWA) operator (Torra, 1997). The WOWA operator generalises both the weighted mean and the OWA operator (Yager, 1988). WOWA can weight values not only according to their importance (like in the weighted mean) but also according to their relative position in the preference scale used. This allows us to define different aggregation strategies depending on the application domain. For instance, in our work, we defined an aggregation function that gives more importance to values closer to extreme values (e.g. +1 and –1) rather than central ones (e.g. +0.1 and –0.1); this implies that users having stronger opinions count more in the decision of accepting or rejecting an image. Our contribution to this research topic consists in defining several decision making criteria and we compare different preference aggregation operators w.r.t. their capability of classifying user behaviour (see Section 7).

Concerning argumentation, some of the approaches that relate to our work are those in argumentation-based decision making (Amgoud and Prade, 2009) and argument-based negotiation (Amgoud et al., 2006). Amgoud and Prade, 2009 proposes a unified argumentation framework for decision making under uncertainty and multiple-criteria decision making that uses arguments to explain the decisions made. Amgoud et al., 2006 defines a negotiation dialogue according to which several agents exchange arguments in order to try to reach an agreement. In these works, an argument is a logical proof for a statement where the proof is built from a knowledge base (containing uncertain information) and a preference base. In our application, on the other hand, arguments are reasons for accepting or rejecting a given image and are essentially tags created by the users. The use of this kind of arguments supports, similarly to the logical approaches mentioned above, the creation of a sense of discourse around a decision since the arguments pinpoint the reasons why users decide to accept or discard a certain image.

As far as the environment enacting the distributed curation workflow and controlling the agents is concerned, our p2p EI infrastructure is based and extends our previous development efforts on engineering multi-agent systems as open agent environments (Noriega, 1997; Esteva, 2003; d'Inverno et al., 2012). Remarkably, we superseded the original conception of d'Inverno et

al., 2012 as centralised systems by an EI infrastructure implemented as a p2p network of nodes that allows us to exploit the benefits inherent to p2p systems (e.g. self-organisation, resilience to faults and attacks, low barrier to deployment, and privacy management). The p2p EI infrastructure used in WeCurate has been developed with the ambition that EIs become a pervasive mechanism to coordinate very large networks of humans and devices in the next years. Our current efforts in improving the infrastructure and a roadmap on EIs development in the last 20 years are reported in de Jonge et al. (2015).

Several systems exist that enable realtime personalised experience and multiuser collaboration in virtual workspaces, both in industry and in academia. In industry, web conferencing software such as Adobe Connect allows complex media and text driven interactions; shared document editors such as Google Drive enable co-editing of office-type documents. However, the user interfaces are perhaps too complex for a casual user in a museum and it is not possible to enforce specific workflows with specific goals with these systems as required by our group curation scenario. Further, agreement technologies such as group decision making are not explicitly supported, e.g. consider the scenario where users are co-editing a presentation using Google Drive and they need to select an appropriate image.

In academia, enhancing the users' experience in museums has already been addressed in different ways. For instance, Dini et al., 2007 outline a multiuser game played on distributed displays. Users are given a mobile device for individual game play, but with situated displays for synchronized public views of shared game play. Therefore, this system is not truly multiuser as they play individually, and the outcome contributes to a shared game. In the PEACH project, researchers focused on the creation of online personalised presentations to be delivered to the visitors for improving their satisfaction and personalised visit summary reports of suggestions for future visits (Kuflik et al., 2005). Their focus was mainly the modeling of preferences of single users but the importance of social interactions in visiting a museum was investigated in the PIL project, an extension of the research results of the PEACH project, and in the ARCHIE project (Kuflik et al., 2011; Luyten et al., 2006). ARCHIE aimed at providing a more socially-aware experience to users visiting a museum by allowing visitors to interact with other visitors by means of their mobile guides. User profiles were used to tailor the information to the needs and interests of each individual user and, as such, no group decision making was necessary. A cultural heritage application was proposed in Costantini et al. (2008) where agents are able to discover users' movements via a satellite, to learn and to adapt user profiles to assist users during their visits in Villa Adriana, an archaeological site in Tivoli, Italy.

## 10. Conclusion and future works

A multiuser museum interactive which uses a multiagent system to support community interactions and decision making and a peer-to-peer Electronic Institution (EI) to model the workflow has been described. Its multimodal user interface which directly represents the scenes in the underlying EI and which is designed to engage casual users in a social discourse around museum artefacts has also been described. An analysis has been presented which assessed the success of the system as a museum interactive as well as the evaluation of various group decision making algorithms implemented in the system.

This line of research looks promising. Our results have shown that the representations of the opinions of the group did influence individual members opinion, which denotes a sense of social presence via the system. The evaluation of decision making

models showed that simple decision making models can predict user behaviour in terms of image collected in a fair way, especially, if we consider that the decision models were based on few activities such as image preferences, zooming, tagging and chatting. We think that these results reveal that the use of agent and EI technology together can enhance user social dynamics and user social presence. This is an important result.

In terms of future work, we can improve user social engagement, the scene design and, the efficacy of the agent architecture in supporting the curation task. For instance, by letting users be more engage in the discussion of images by taking advantage of *gamification* in the design of the forum scene.

Another interesting extension of the system is the allowance of more complex arguments, alluding expert opinions, similar past opinions or value-based opinions, etc. On the one hand, having these kinds of more complex argument structures can foster the modeling of more advanced decision making models and, consequently, the development of a more elaborated analysis of the agents' behaviour. On the other hand, they will likely require a new GUI design for maintaining the usability of the interface, a key element for conveying a sense of shared experience to the users of our system. In fact, the challenge lies more on maintaining an intuitive user interface for the novice user, than increasing the complexity of the argumentation framework.

We also wish to revisit an idea that was in our earlier prototype (Yee-King et al., 2012), where an online image recommender was used to select images that matched the tag preferences of two users. The idea of recommending images in this way was rejected for the WeCurate system after several users reported frustration at receiving a series of similar images (Hazelden et al., 2012). A smarter method would be to extract a representation of images based on their potential for discussion by the group, as opposed to a simplistic, tag based metric. For example, which parts of the images were users zooming into? Which types of image engendered the most active discussion?

Beyond that, the technology has been designed to easily transfer to a web or mobile application, and the distributed peer-to-peer Electronic Institution model is designed to scale; and we see great potential in the concept of agent supported, workflow driven, synchronous image discussion and curation taken to the mass audience on the open web. This paper contributes to the integration of agent-based and human-based decision making processes in socio-technical systems. We consider this a key research area in the design of intelligent agents.

## Acknowledgements

The authors would like to thank the reviewers for their valuable comments. This work is supported by the Collective Mind project (Spanish Ministry of Economy and Competitiveness, under Grant number TEC2013-49430-EXP), the PRAISE project (funded by the European Commission, under Grant number 388770), and the ACE project (funded by the European framework ERA-Net CHIST-ERA, under contract CHRI-001-03). Roberto Confalonieri would like to thank Andrea Mattivi for his suggestions on the data analysis.

## References

Albert, C., Amgoud, L., Costedoat, C., Bannay, D., Saint-Dizier, P., 2011. Introducing argumentation in opinion analysis: language and reasoning challenges. In: *Proceedings of Sentiment Analysis Where AI Meets Psychology* (workshop at IJCNLP11).

Amgoud, L., Belabbès, S., Prade, H., 2006. Towards a formal framework for the search of a consensus between autonomous agents. In: Parsons, S., Maudet, N.,

Moraitis, P., Rahwan, I. (Eds.), *Argumentation in Multi-Agent Systems*, Lecture Notes in Computer Science, vol. 4049, Springer, Berlin, Heidelberg, pp. 264–278.

Amgoud, L., Dimopoulos, Y., Moraitis, P., 2007. A unified and general framework for argumentation-based negotiation. In: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, ACM Press, Honolulu, Hawaii, USA, pp. 963–970.

Amgoud, L., Prade, H., 2009. Using arguments for making and explaining decisions. *Artif. Intell. J.* 173, 413–436.

Amgoud, L., Besnard, P., 2009. Bridging the gap between abstract argumentation systems and logic. In: *Proceedings of International Conference on Scalable Uncertainty Management (SUM'09)*, pp. 12–27.

Amgoud, L., Confalonieri, R., Jonge, D.D., d'Inverno, M., Hazelden, K., Osman, N., Prade, H., Sierra, C., Yee-King, M., 2012. Sharing online cultural experiences: an argument-based approach. In: Torra, V., Narukawa, Y., López, B., Villaret, M. (Eds.), *Modeling Decisions for Artificial Intelligence – 9th International Conference, MDAI 2012, Girona, Catalonia, Spain, Lecture Notes in Computer Science*, vol. 7647, Springer, pp. 282–293.

Arcos, J.L., Esteve, M., Noriega, P., Rodríguez-Aguilar, J.A., Sierra, C., 2005. Engineering open environments with electronic institutions. *Eng. Appl. Artif. Intell.* 18 (2), 191–204.

Atkinson, K., Bench-Capon, T., McBurney, P., 2004. Justifying practical reasoning. In: *Proceedings of the 4th Workshop on Computational Models of Natural Argument (CMNA'04)*, pp. 87–90.

Barron, C., 2012. The Washington Post, The Art of Video Games.

Benferhat, S., Dubois, D., Kaci, S., Prade, H., 2006. Bipolar possibility theory in preference modeling: representation, fusion and optimal solutions. *Inf. Fusion* 7 (1), 135–150.

Bonet, B., Geffner, H., 1996. Arguing for decisions: a qualitative model of decision making. In: *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI'96)*, pp. 98–105.

Cosley, D., Lewenstein, J., Herman, A., Holloway, J., Baxter, J., Nomura, S., Boehner, K., Gay, G., 2008. Artlinks: fostering social awareness and reflection in museums. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*, ACM, New York, NY, USA, pp. 403–412.

Cosley, D., Baxter, J., Lee, S., Alson, B., Nomura, S., Adams, P., Sarabu, C., Gay, G., 2009. A tag in the hand: supporting semantic, social, and spatial navigation in museums. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*, ACM, New York, NY, USA, pp. 1953–1962.

Costantini, S., Mostarda, L., Tocchio, A., Tsintza, P., 2008. DALICA: agent-based ambient intelligence for cultural-heritage scenarios. *IEEE Intell. Syst.* 23 (2), 34–41.

d'Inverno, M., Luck, M., Noriega, P., Rodríguez-Aguilar, J.A., Sierra, C., 2012. Communicating open systems. *Artif. Intell.* 186, 38–94.

de Jonge, D., Rodríguez-Aguilar, J.A., Rosell, B., Sierra, C., 2014. Infrastructures to engineer open environments as electronic institutions, IFAAMAS, IFAAMAS, Paris, France.

de Jonge, D.D., Rosell, B., Sierra, C., 2013. Human interactions in electronic institutions. In: Chesñevar, C.I., Onaindia, E., Ossowski, S., Vouras, G.A. (Eds.), *Agreement Technologies – Proceedings of Second International Conference, AT 2013, Beijing, China, August 1–2, 2013, Lecture Notes in Computer Science*, vol. 8068, Springer, pp. 75–89.

de Jonge, D., Rodríguez-Aguilar, J.A., Rosell, B., Sierra, C., 2015. Infrastructures to Engineer Open Environments as Electronic Institutions. In *Agent Environments for Multi-Agent Systems*, Springer, in press.

Dini, R., Paternò, F., Santoro, C., 2007. An environment to support multi-user interaction and cooperation for improving museum visits through games. In: *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'07)*, MobileHCI'07, ACM, New York, NY, USA, pp. 515–521.

Dung, P.M., 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell. J.* 77, 321–357.

Esteve, M., de la Cruz, D., Sierra, C., 2002. ISLANDER: an electronic institutions editor. In: *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems: Part 3, AAMAS'02*, ACM, New York, NY, USA, pp. 1045–1052.

Esteve, M., 2003. *Electronic Institutions: from Specification to Development*. (Ph.D. thesis), Technical University of Catalonia.

Gaertner, D., García-Camino, A., Noriega, P., Rodríguez-Aguilar, J.-A., Vasconcelos, W., 2007. Distributed norm management in regulated multiagent systems. In: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS '07*, ACM, New York, NY, USA, pp. 90:1–90:8. <http://dx.doi.org/10.1145/1329125.1329235>.

Gaitatzes, A., Roussou, M., 2002. Reviving the past: cultural heritage meets. In: *Proceedings of the 2001 Conference on Virtual reality, Archeology, and Cultural Heritage*, pp. 103–110.

Glaser, B., Strauss, A., 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine De Gruyter, New York, NY.

Glaser, B., 2008. *Doing Quantitative Grounded Theory*. Sociology Press.

Greg, A., Your paintings: public access and public tagging. *J. Scott. Soc. Art Hist.* 16, 2011.

Hazelden, K., Yee-King, M., Amgoud, L., d'Inverno, M., Sierra, C., Osman, N., Confalonieri, R., Jonge, D.D., 2012. WeCurate: designing for synchronised browsing and social negotiation. In: Ossowski, S., Toni, F., Vouras, G.A. (Eds.), *Proceedings*

- of the 1st International Conference on Agreement Technologies, AT 2012, CEUR Workshop Proceedings, CEUR-WS.org, vol. 918, pp. 168–179.
- Hazelden, K., Yee-King, M., Confalonieri, R., Sierra, C., Ghedini, F., Jonge, D.D., Osman, N., d'Inverno, M., 2013. WeCurate: multiuser museum interactives for shared cultural experiences. In: Mackay, W.E., Brewster, S.A., Bødker S. (Eds.), 2013 ACM SIGCHI Conference on Human Factors in Computing Systems, CHI'13, Paris, France, April 27 – May 2, 2013, Extended Abstracts, ACM, pp. 571–576.
- Heath, C., Lehn, D.V., Osborne, J., 2005. Interaction and interactives: collaboration and participation with computer-based exhibits. *Public Underst. Sci.* 14 (1), 91–101.
- Hellin-Hobbs, Y., 2010. The constructivist museum and the web. In: Seal, A., Bowen, J.P., Ng, K. (Eds.), Proceedings of the 2010 International Conference on Electronic Visualisation and the Arts (EVA'10), British Computer Society, Swinton, UK, UK, pp. 72–78.
- Hope, T., Nakamura, Y., Takahashi, T., Nobayashi, A., Fukuoka, S., Hamasaki, M., Nishimura, T., 2009. Familial collaborations in a museum. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09), ACM, New York, NY, USA, pp. 1963–1972.
- Hornecker, E., 2008. I don't understand it either, but it is cool—visitor interactions with a multi-touch table in a museum. In: 2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems, pp. 113–120.
- Krishna-Bal, B., Saint-Dizier, P., 2010. Towards building annotated resources for analyzing opinions and argumentation in news editorials. In: Proceedings of the International Conference on Language Resources and Evaluation.
- Kuflik, T., Callaway, C., Goren-Bar, D., Rocchi, C., Stock, O., Zancanaro, M., 2005. Non-intrusive user modeling for a multimedia museum visitors guide system. In: Ardissono, L., Brna, P., Mitrovic, A. (Eds.), Proceedings of the 10th International Conference on User Modeling, Lecture Notes in Computer Science, vol. 3538. Springer-Verlag, Berlin, Heidelberg, pp. 236–240.
- Kuflik, T., Stock, O., Zancanaro, M., Gorfinkel, A., Jbara, S., Kats, S., Sheidin, J., Kashtan, N., 2011. A visitor's guide in an active museum: presentations, communications, and reflection. *J. Comput. Cult. Herit.* 3 (3), 11:1–11:25.
- Luyten, K., van Loon, H., Teunkens, D., Gabriels, K., Coninx, K., Manshoven, E., 2006. ARCHIE: disclosing a museum by a socially-aware mobile guide. In: Proceedings of the 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2006).
- Modgil, S., Faci, N., Meneguzzi, F., Oren, N., Miles, S., Luck M., 2009. A framework for monitoring agent-based normative systems. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '09, International Foundation for Autonomous Agents and Multiagent Systems, vol. 1. Richland, SC, pp. 153–160. (<http://www.dl.acm.org/citation.cfm?id=1558013.1558034>).
- Noriega, P., 1997. Agent Mediated Auctions: the Fishmarket Metaphor. (Ph.D. thesis), Autonomous University of Barcelona.
- Pollock, J., 1992. How to reason defeasibly. *Artif. Intell. J.* 57, 1–42.
- Osman, N., Sierra, C., Sabater, J., 2006. Propagation of opinions in structural graphs. In: Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'2006).
- Osman, N., Sierra, C., Mcneill, F., Pane, J., Debenham, J., 2014. Trust and matching algorithms for selecting suitable agents. *ACM Trans. Intell. Syst. Technol.* 5 (1), 16:1–16:39.
- Prakken, H., 2005. Coherence and flexibility in dialogue games for argumentation. *J. Log. Comput.* 15, 1009–1040.
- Ribeiro, R.A., 1996. Fuzzy multiple attribute decision making: a review and new preference elicitation techniques. *Fuzzy Sets Syst.* 78 (2), 155–181.
- Rowstron, A.I.T., Druschel, P., 2001. Pastry: scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In: Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms Heidelberg, Middleware '01, Springer-Verlag, London, UK, UK, pp. 329–350.
- Subrahmanian, V., 2009. Mining online opinions. *IEEE Comput.*, 88–90.
- Torra, V., 1997. The weighted owa operator. *Int. J. Intell. Syst.* 12 (2), 153–166.
- Turner, C., 2011. Making native space: cultural politics, historical narrative, and community curation at the national museum of the american indian. *Pract. Anthr.* 33 (2), 40–44.
- Walker, K., 2008. Digital technologies and the museum experience. Handheld guides and other media, Altamira Press—Rowman & Littlefield Publishers, Inc., Ch. Structuring visitor participation.
- Wojciechowski, R., Walczak, K., White, M., Cellary W., 2004. Building virtual and augmented reality museum exhibitions. In: Proceedings of the 9th International Conference on 3D Web Technology, pp. 135–144.
- Yager, R.R., 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* 18 (1), 183–190.
- Yee-King, M., Confalonieri, R., Jonge, D.D., Osman, N., Hazelden, K., Amgoud, L., Prade, H., Sierra, C., d'Inverno, M., 2012. Towards community browsing for shared experiences: the WeBrowse system. In: Ossowski, S., Toni, F., Vouros, G. A. (Eds.), Proceedings of the First International Conference on Agreement Technologies, AT 2012, CEUR Workshop Proceedings, CEUR-WS.org, vol. 918, pp. 201–202.
- Yee-King, M., Confalonieri, R., Jonge, D.D., Hazelden, K., Sierra, C., d'Inverno, M., Amgoud, L., Osman, N., 2013. Multiuser museum interactives for shared cultural experiences: an agent-based approach. In: Ito, Jonker, Gini, Shehory (Eds.), Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), ACM, pp. 917–924.