What if gamified software is fully proactive? Towards autonomy-related design principles *

Esteban Guerrero[1[0000−0002−6035−800X]], Tero Vartiainen[1[0000−0003−3843−8561]], and Panu Kalmi[2[0000−0002−1359−5109]]

[1] School of Technology and Innovations, University of Vaasa, Finland
[2] School of Accounting and Finance, Economics, University of Vaasa, Finland
{esteban.guerrero, tero.vartiainen, panu.kalmi}@uwasa.fi

Abstract. Computational agents are a type of software architectures designed to be *autonomous* and *social*, meaning that they can make decisions proactively, reacting also to stimuli from the environment. The use of such architectures is not common in the *gamification* field area, instead, gamified software has traditionally reactive characteristics, responding to user actions disregarding the possibility of proactive behavior. In this paper, we propose four formal principles for designing autonomous gamified systems, to ensure *traceability* of gamified outputs, *internal consistency* of gamification attempts, *coherent* agent-user interaction, and formal conditions to assess user actions from a *rational* perspective. We present our initial work on these general principles, highlighting our empirical future work.

Keywords: Persuasive technology · Gamification · Software agents · Principles · Formal approaches · Argumentation dialogues.

1 Motivation

In the artificial intelligence (AI) field, the sense of "*autonomy*" is not precise, but the term is taken to mean that *software agents'* activities do not require constant human guidance or intervention [19]. An object is an *agent* (*e.g.* a software agent) if it serves a useful purpose either to a different agent, or to itself, in which case the agent is called *autonomous* [12]. Being those purposes the state of affairs to be achieved, in other words, the *goals* of an agent. To exemplify this notion, let us suppose that a person wants to increase her monthly savings by counting the accumulated expenses using a *financial app*. That software is her agent that keeps her savings count, "adopting" her goal to motivate herself for saving money. Note that the app has a *transient* agency, if the user does no longer needs to save money (*e.g.* she wins the lottery), such an agent becomes a simple *object* with no *ascribed* agency.

* Persuasive 2022, Adjunct Proceedings of the 17th International Conference on Persuasive Technology.

An *autonomous agent* is not dependent on the goals of others, it possesses goals that are *generated* from within rather than adopted [20]. For example, an autonomous version of that financial app, could change its goal (proactively) and help her to visualize different philanthropic goals, without any guidance or intervention.

A high-level question directing this research connects the aforementioned autonomous agents and gamification field, *what if a gamified software became fully autonomous?*. Empirical answers to this question from a gamification perspective are scarce, and theoretical frameworks of gamification dealing with this issue are practically non-existent (see reviews [6,11]). From the AI perspective, "human oversight" is one of the requirements being put forward as a means to support human autonomy and agency [13], where theoretical (we will use the term *formal*) guidelines have been proposed, aiming to delineate autonomous behavior of agents considering responsible and transparent mechanisms [2]. In fact, high-level principles and guidelines [8,14,16] are commonly used in gamification, but most of them are not *aligned* with autonomous software characteristics, nor serve as grounded specifications for developing actual software. In this context, our ongoing research proposes four principles for designing autonomous gamification technology considering: *traceability* as a mechanism for meaningful human control, *coherence* and *consistency* during the interaction autonomous agent-human, and *rationality* of the decisions that a proactive agent makes. In summary, the proposed principles are:

Principle 1 : *Traceability of gamified outputs.* Establishes that gamified affordances (outputs) need to provide a transparent and identifiable explanation of the persuasive attempt.

Principle 2 : *Internal consistency of a gamification attempt.* Defines formal requirements of the informational elements (*e.g.* content, visualization, etc.) of a persuasive attempt.

Principle 3 : *Coherent gamified interaction.* Characterizes the type of interaction that a persuasive agent should (or should not) make.

Principle 4 : *Rational persuasive gamification*: Determine the formal conditions that an agent needs to consider if a user action can be considered as *rational.*

We formalize these principles in *propositional logic* to be used by designers of formal mechanisms of decision-making for software agents (*e.g.* [4] among others).

2 Methods

We use a formal method (framework) based on *argumentation-based games* [17] for describing the agent-user interaction. In this paper, a user model is a tuple: $\mathcal{U} = \langle \mathcal{B}, \mathcal{I}, \mathcal{B}e, \mathcal{P}_{\mathcal{I}}, \mathcal{P}_{\mathcal{B}}, \preceq_{\mathcal{B}}, \preceq_{\mathcal{I}} \rangle$. In which the probability distributions $\mathcal{P}_{\mathcal{I}}$ and $\mathcal{P}_{\mathcal{B}}$ relate the subjective probability of intentions and beliefs, and hierarchies of beliefs and intentions are given by $\preceq_{\mathcal{B}}, \preceq_{\mathcal{I}}$.

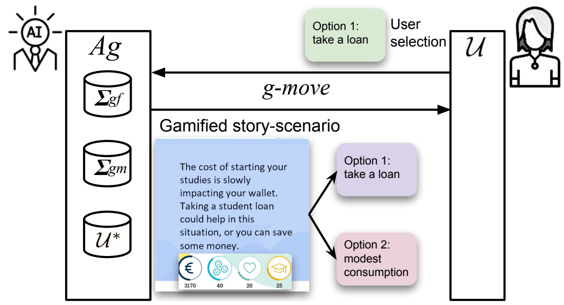

Fig. 1: The *user-agent* gamified interaction used in this paper.

In this paper, we consider *gamification mechanisms* (gm) *e.g.* avatars, stories and leader boards, and *gamification feedback* (gf), which are visual *affordances* presented to a user *e.g.* rewards, feedback messages, etc. This classification of gamification affordances follows a taxonomy presented in [5]. We assume that the agent has two *databases* Σ_{gm} and Σ_{gf}, containing gm and gf affordances. We also consider preferences among affordances gm and gf, which is a pre-order function $\preceq_{gm}$ and $\preceq_{gf}$, we assume a preexisting order. In this context, a user and an agent exchange information regarding a particular *topic* $\mathcal{T}$, *e.g.* about financial literacy.

We use *propositional logic* with $\neg$ to express logical negation, $\bar{x}$ denoting uncertainty (*w.r.t.* a *true*, *false* valuation), $\vdash$ deductive inference, and $\vdash_s$ semantic interpretation, and $\equiv$, $\equiv_s$ for syntactic and semantic equivalence. We also use a handy function for updating information $\mathsf{UPD}(old, new)$. Agent Ag, as a gamified persuasive technology is oriented to generate as *output* a ***gamified move*** (*gmove*), which is a tuple $\langle sa, cont, vis\rangle$ formed by: 1) a *speech act* (*sa*) that is the intended action of the agent within the persuasive exchange as a dialogue, 2) a *persuasive content* (*cont*) that is the underlying message to be transmitted to a user, and 3) a *visual cue* (vis). *sa* are predefined actions such as *accept, assert, question, reject, ignore* (see technical details in [5]). We use the notation $gmove_{t_i}^{Ag\to\mathcal{U}}$ to express that a *gmove* was made from Ag to $\mathcal{U}$ at t_i, which is omitted if time and move direction is evident or unnecessary. We also use three handy functions $\mathsf{CONT}(gmove) = \{content\}$, $\mathsf{VIS}(gmove) = \{visual\}$ and $\mathsf{SA}(gmove) = \{$ *speech act* $\}$ to return the content and the visualization of a given *gmove*. An agent Ag uses as *input* the aforementioned three databases Σ_T, Σ_{gf} and Σ_{gm}, and a ***model of the user*** $\mathcal{U}^*$ (where $\mathcal{U}^* \subseteq \mathcal{U}$, denoting that it may have not perfect information).

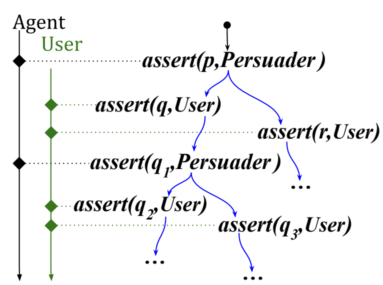

Fig. 2: Protocol of gamified persuasive interaction

3 Results - Principles for autonomous gamified systems

We define these principles considering three assumptions: 1) a user establishes a communication with the agent using gamified affordances, 2) the agent has information about beliefs, intentions, and preferences of the user, and 3) the user follows some protocols of communication with the agent.

Principle 1 (Traceable gamified output) *A persuader agent should be able to provide traceable explanations for every gamified output. Formally, if* $S \vdash_s gmove^{Ag \rightarrow \mathcal{U}}$ *is the gamified move output, and* S *is the knowledge source of the move, the following criteria should be fulfilled:*

Formalism	***Explanation***
$S \subseteq (\Sigma_{gf} \cup \Sigma_{gm} \cup \mathcal{U}^*)$	*A gamified output of an agent should be the consequence of an inference process based on a set of gamified mechanisms and the user model.*
$S \neq \emptyset$	*Determinants of a gamified move should be identifiable.*

This first principle relies on the traceability of the semantic inference ($\vdash_s$). In the AI literature, violations of these formalisms have been investigated to avoid *black-box*-style algorithms [18], and *handcrafted* processes with no *back-track* inference mechanism.

Focusing on the internal definition of the gamified output, the following set of principles establishes basic conditions of consistency of such output.

Principle 2 (Internal consistency of a gamification move) *A gamification move is consistent if the following holds:*

Formalism	***Explanation***
$\mathsf{CONT}(gmove^{Ag \rightarrow \mathcal{U}}) \equiv_s \mathsf{VIS}(gmove^{Ag \rightarrow \mathcal{U}})$	*The content and the visualization of a gamified move should be semantically coherent.*
$\mathsf{CONT}(gmove^{Ag \rightarrow \mathcal{U}}) \subseteq \mathcal{T}$	*The content of a gamified move should be within the agreed persuasive topic.*
$\mathsf{VIS}(gmove^{Ag \rightarrow \mathcal{U}}) \in \preceq_{gm} \cup \preceq_{gf}$	*The visualization should be part of an agreed set of gamified affordances.*

Principle 2 can be seen close as a design principle for visual and content aspects of a gamification process, in which the information carried by a gamified affordance has to be consistent with its visualization.

The next set of principles is related to the gamified interaction between agent and user, specifically, these formalisms try to avoid design practices against engagement and commitment of the agent to the users' goals.

Principle 3 (Coherent gamified interaction) *A gamified persuasive interaction between* $\mathcal{U}$ *and* Ag *is coherent if the following conditions hold:*

Formalism	*Explanation*
$gmove_{t_i} \neq gmove_{t_{i+1}}$	*A gamified persuasive output should not be* ***repetitive*** *regardless of the state of* $\mathcal{U}$
$\forall\ gmove_{t_i}^{\mathcal{U}\rightarrow Ag}$, $\mathsf{SA}(gmove_{t_{i+1}}^{\mathcal{U}\rightarrow Ag}) \setminus \{ignore\}$	*A persuasive agent* ***should not ignore*** *petitions from a user.*
$\mathsf{SA}(gmove^{\mathcal{U}\rightarrow Ag}) = \{ignore\}$, *then* $\mathsf{UPD}(\preceq_{gf}, \preceq_{gf})$	*When a user* ***ignores*** *a gamification move, then the agent must update the gamification feedback preferences.*
$\mathsf{SA}(gmove^{\mathcal{U}\rightarrow Ag}) = \{reject\}$, *then* $\mathsf{UPD}(\preceq_{gf}, \preceq_{gf})$, $\mathsf{UPD}(\mathcal{U})$	*When a user* ***rejects*** *a gamification move, the agent must update the gamification feedback preferences and the beliefs of the user model.*

The set of principles defined as coherent gamified interactions establishes a guideline checklist of what an agent should do when a user communicates directly through gamified moves.

A last set of principles establishes conditions to assess the *rational* actions of a user. In the agents' literature, the user's *intention* to X (*e.g.* $X = walk$) provides the agent with support to *believe* that s/he will do X, *i.e.* beliefs and intentions are in the same "direction" [3]. The following formalisms capture the notion of rationality considering an *alignment* between beliefs, intentions, and actions.

Principle 4 (Rational persuasive gamification) *An agent Ag persuasive gamification can be considered rational, incomplete or irrational if the following holds:*

Formalism	*Explanation*
If $\mathsf{CONT}(gmove^{\mathcal{U}\rightarrow Ag}) \in \mathcal{B}$ *and* $\preceq_{\mathcal{I}}$, *then such move is rational*	*A move is* ***rational*** *if it contains a belief that is in the agent's user model, and it is part of the hierarchy of preferred intentions.*
If $\mathsf{CONT}(gmove^{\mathcal{U}\rightarrow Ag}) \in \preceq_{\mathcal{I}}$ *and* $\overline{\mathsf{CONT}(gmove^{\mathcal{U}\rightarrow Ag})} \in \mathcal{B}$, *then* $\mathcal{U}^*$ *is intention-belief-incomplete*	*The model of a user is* ***intention-belief-incomplete*** *if a move is inline with the preferred intentions but contrary to the belief model.*
If $\mathsf{CONT}(gmove^{\mathcal{U}\rightarrow Ag}) \in \preceq_{\mathcal{I}}$ *and* $\neg(\mathsf{CONT}(gmove^{\mathcal{U}\rightarrow Ag})) \in \mathcal{B}$, *then such move is* ***irrational***	*A move is irrational if the move content is part of the hierarchy of preferred intentions but at the same time is against the set of beliefs.*

4 Discussion and future work

Current gamification literature (see reviews [10,11]) shows five main trends: 1) the popularity of a limited number of affordances such as *leader-boards*, rewarded points, and textual or visual feedback; 2) goal-orientation of the theoretical gamification foundation; 3) the extended use of competitiveness and cooperativeness mechanisms for gamification, and 4) a gradual generalization of tailoring gamified persuasion. However, most of these approaches do not consider the notion of autonomy or proactiveness of the gamified software.

On the AI-side, persuasion is a well-established research track, especially in the *argumentation theory* [9]. Computational persuasion is a highly regulated process that leads to the design of *argumentation-based dialogue games*, which is a *protocol*-based exchange of information between two agents. Shortcomings of these dialogues are well-identified (see [9]), such as the harnessing of finding *optimal* decisions (*moves*) at every game stage, and the use of only *exocentric persuasion* [7] disregarding context information or mental states of a user.

We introduced a set of rules aiming to promote *transparency* of persuasive attempts (Principle 1), which are in line with the current joint effort that the European Union and other leading AI countries to develop *guidelines for trustworthy AI*[3]. In the human-computer interaction field, *trustworthiness* was highlighted to be a fundamental principle for *system credibility* [16], explainable and persuasive interfaces [15]. Consistency between visual and content aspects is relevant for most of the gamification used in persuasive attempts. Principle 2 is linked with the work presented in [14], where Nemery et al. have highlighted the importance of visual consistency, introduced a set of principles for persuasive interfaces. The third set of formalisms, Principle 3, establishes a minimum set of rules for coherent interaction between an agent and a user. We acknowledge that the type of interactions are linked to argumentation-based dialogues, which limits the type of potential interactions that other gamification mechanisms can produce. Nevertheless, Principle 3 is a generalization of different proactive gamification mechanisms that have been investigated in the literature (see [5]). Finally, our set of principles for evaluating rational inputs from the user (Principle 4), which are based on the well-established theory of *practical reason* of Bratman in [1], establishes a guide for evaluating if the user's actions are aligned with the information that an agent possesses about the user.

A key **limitation** of this work is the lack of empirical validation. Our immediate future work will be to evaluate empirically these principles using a real-world scenario. Currently, we are designing the following version of the gamified platform (omitted details for blind review) to support financial decisions. We also want to further establish an axiomatization of those principles, considering different types of gamification affordances and different types of software agents (*e.g.* [5]).

References

1. Bratman, M.: Intention, plans, and practical reason. Harvard University Press (1987)
2. Dignum, V.: Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer International Publishing (2019)
3. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Addison-Wesley series in social psychology, Addison-Wesley Publishing Company (1975)

[3] *Ethics guidelines for trustworthy AI* see https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, last access November 29 2021

4. Guerrero, E., Lindgren, H.: Practical Reasoning About Complex Activities. SpringerLink pp. 82–94 (Jun 2017)
5. Guerrero, E., Lindgren, H.: Typologies of persuasive strategies and content: a formalization using argumentation. In: International Conference on Practical Applications of Agents and Multi-Agent Systems. pp. 101–113. Springer (2021)
6. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work?–a literature review of empirical studies on gamification. In: 2014 47th Hawaii international conference on system sciences. pp. 3025–3034. Ieee (2014)
7. de la Hera Conde-Pumpido, T.: Persuasive gaming: Identifying the different types of persuasion through games. International Journal of Serious Games **4**(1), 31–39 (2017)
8. Horvitz, E.: Principles of mixed-initiative user interfaces. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. pp. 159–166 (1999)
9. Hunter, A.: Computational Persuasion with Applications in Behaviour Change. In: COMMA. pp. 5–18 (2016)
10. Klock, A.C.T., Gasparini, I., Pimenta, M.S., Hamari, J.: Tailored gamification: A review of literature. International Journal of Human-Computer Studies **144**, 102495 (2020)
11. Koivisto, J., Hamari, J.: The rise of motivational information systems: A review of gamification research. International Journal of Information Management **45**, 191–210 (2019)
12. Luck, M., d'Inverno, M., et al.: A formal framework for agency and autonomy. In: Icmas. vol. 95, pp. 254–260 (1995)
13. Methnani, L., Aler Tubella, A., Dignum, V., Theodorou, A.: Let Me Take Over: Variable Autonomy for Meaningful Human Control. Front. Artif. Intell. **0** (2021)
14. Némery, A., Brangier, E.: Set of guidelines for persuasive interfaces: Organization and validation of the criteria. Journal of Usability Studies **9**(3) (2014)
15. Némery, A., Brangier, E., Kopp, S.: First Validation of Persuasive Criteria for Designing and Evaluating the Social Influence of User Interfaces: Justification of a Guideline. In: Design, User Experience, and Usability. Theory, Methods, Tools and Practice, pp. 616–624. Springer, Berlin, Germany (Jul 2011)
16. Oinas-Kukkonen, H., Harjumaa, M.: Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems **24**(1), 28 (2009)
17. Parsons, S., Wooldridge, M., Amgoud, L.: An analysis of formal inter-agent dialogues. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1. pp. 394–401. ACM (2002)
18. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead - Nature Machine Intelligence. Nat. Mach. Intell. **1**, 206–215 (May 2019)
19. Shoham, Y.: Agent-oriented programming. Artificial intelligence **60**(1), 51–92 (1993)
20. Wooldridge, M., Jennings, N.R.: Agent theories, architectures, and languages: A survey. SpringerLink pp. 1–39 (Aug 1994)