

Designing a Creature Believability Scale for Videogames

Nuno Barreto^(✉), Rui Craveirinha, and Licinio Roque

CISUC, Department of Informatics Engineering, University of Coimbra,
3004 516 Coimbra, Portugal
nbarreto@dei.uc.pt

Abstract. This paper describes the design, and early evaluation of a scale aimed at assessing the believability of creatures in videogames. These creatures include all zoomorphic entities that do not qualify as fundamentally human-like, whether or not they have characteristics identifiable as anthropomorphic. The work is based on principles drawn from biology, animation, illustration and artificial intelligence. After developing the scale's 46 original items, it was administrated as a Likert Scale questionnaire. The results were analyzed through Principal Component Analysis and they suggest that 26 items, out of the original 46, spread across 4 dimensions, could be used to evaluate creature believability.

Keywords: Believability scale · Game design · Experience evaluation · Videogames

1 Introduction

This paper reports on the process of creation of a scale to assess believability of a subset of videogame creatures. We limit this subset to zoomorphic entities, inspired by (contemporary or otherwise) living or fictional beings, that are not identifiable as human, nor fundamentally human-like, despite whatever anthropomorphic characteristics they might have. Creatures belonging to this subset will henceforth be referred to as creatures, for simplicity.

Defined, by Fogg and Tseng, as a synonym for credibility, believability is a construct measured by the perception users have of a given artifact [8]. This is because the authors consider artifacts as a source of information and their perceived veracity contributes to their believability. In multimedia, or at least in videogames, believability can be fundamental in inducing immersion [39], a form of Presence [18], more so when such artifacts deal with unrealistic themes (fantasy, sci-fi, mythology, etc.).

In particular, believability can be a crucial element of the game's world's actors, leading humans to accept they are "alive and thinking" [28]. This is the goal behind believable agents, an artificial intelligence field focused on simulating agents with credible behaviors. However, developing believable agents transcends

artificial intelligence as it is assumed to be one of the driving factors when crafting videogame experiences [39]. In fact, studies have shown such actors are perceived to be more engaging than non-believable ones [2, 13, 28, 39].

Studies in believability mainly follow two directions: one studying how to create behavioral patterns which, when observed, incite believability. This is, as previously explained, the scope of the Believable Agents field study [13]. The other, has its roots in animation where artists were concerned with creating life-like beings, then known as believable characters. Here, anthropomorphism was key to make viewers relate to them [2]. Nonetheless, the literature behind both directions appears to focus extensively on simulating humans. On one hand, as either in-game characters, or players [36] and, on the other, as compelling and emotional characters, part of a narrative [2]. It is worth noting however, that believability is in no way a synonym for realism. In fact, some of the construct's earlier studies were conducted with cartoons [35]. This strengthens the idea that believability is achieved from the mental model constructed via an observer's interpretation of a given artifact.

While synthetic humans, and humanoids, are abundant in videogames, they are not the only type of existing virtual "living"-entities. The overwhelming focus on the first, despite the existence of the latter, helps set the ground for our research. Specifically, we aim to study believability in creatures, under the hypothesis that, similarly to how studies suggest virtual humans/humanoids are more engaging if considered believable, creatures are also more appealing if presented in a believable manner. This is particularly important on games where they either play a central role, such as Life-Simulation games, or are an integral part in supporting the game's environment (this is most notably true in the fantasy open-world genre).

However, while there is already a well-defined set of expectations towards humans, [2] (which helps their believability assessment), there is, as far as could be assessed, none for creatures. Therefore, in order to verify our hypothesis, we must first answer "What makes a creature feel believable?". The scale proposed in this paper is a step towards that goal. Moreover, by creating such a tool, potential pitfalls in existing videogame creatures may be identified, and the design of future ones may be founded on a better understood basis. The contribution of this paper is then a methodological one [42]: an instrument to measure creature believability and aid the design process to maximize that perception.

This paper is structured as follows: Sect. 2 will describe the methodology behind the scale's construction including item generation and the setup behind administering the scale. This is followed by Sect. 3, Results and Scale Revision, where the data collected is used to validate and help revise the scale itself. Finally, Sect. 4, will conclude the paper.

2 Methodology

This research is being developed using a Design Science Research methodology [12] (focused on the creature believability scale as a model output), with this

paper’s work corresponding to an iteration cycle. Next we will present the following steps in the process: problem awareness, and its foundation on previous literature, construction of the creature believability scale proposal and a first evaluation and revision based on Principal Components Analysis.

2.1 Creature Believability Scale Construction

The main reason behind the creation of a scale, over another instrument, was due to the nature of believability. Because it is a construct originating from an observer’s perception, we chose rating scales [32].

The scales’s proposal construction underwent a three-step process, inspired by the method described by Spector [32]. Firstly, we defined what we would consider “creatures”, establishing a division between humans (and humanoids) and the entities under our study. Based on the definition proposed by Whithlatch [41] (which went in the desired direction) we defined creatures as stated in the introduction: zoomorphic entities, inspired by (contemporary or otherwise) living or fictional beings, that are not identifiable as human, nor fundamentally human-like, despite whatever anthropomorphic characteristics they might, or might not, have.

The second step consisted in defining a list of underlying constraints of our scale:

1. Unlike humans, and humanoids, who have a distinct (limited) set of characteristics, our definition includes markedly heterogeneous beings, ranging from insectoid to mammal-like. With this in mind, our first concern was to create a sufficiently broad set of identifiable elements to evaluate the wide variety of creatures.
2. Instead of considering believability as a binary factor, we opted to work with a fuzzy set. This would allow us to better quantify the qualitative, and variable, nature of the believability a given creature may have, as well as better identify which elements contribute to that perception.
3. As believability is intrinsically perceptual [8], we considered limiting the evaluation of the creatures to perceivable (phenotypic) elements.
4. Whilst the creature definition given includes entities inspired by both living and fictional beings, the everyday experience of a human is with living beings. Therefore, the starting point for the evaluation elements was plausible characteristics.
5. Finally, since our objective is to validate and, subsequently, use this scale with gamers, which may or may not have a scientific background, the language was made accessible, deprived of technical terminology.

After analyzing the constraints, we constructed several candidate statements. This was done both through induction on examples and with the support of extant literature. For the latter we surveyed multiple study fields such as, but not limited to, believable agents, ethology, biology, human perception of living-beings, illustration and animation.

This work began with a survey of existing believable-agents literature to retrieve cognition-related items. One source, by Togelius, et al. [36], cited ConsScale [25] as a basis for their work. ConsScale is an evaluation tool designed to quantify the “level of consciousness” in artificial agents [25] by analyzing their architecture. The tool is composed of 13 levels where each contains a set of statements illustrating the cognitive skills an agent must have at that given level. These levels go from “Disembodied” (entities lacking defined boundaries or cognitive skills) to “Super-Conscious” (entities surpassing human-beings, capable of managing several consciousnesses simultaneously).

Similarly to Togelius, et al., we used ConsScale to derive some of our scale’s items, albeit with some preprocessing. Firstly, we used only a subset of the scale, each of the scale’s levels’ biological phylogeny working as a heuristic to observe constraints 1 and 4. Thus, only level 2 (viruses) through 8 (primates) were used. Subsequently, we rewrote the levels’ items to enumerate the phenotypic traits corresponding to the respective statements. The resulting items accounted for reaction to the environment, intention (underlined by the human need to attribute intention to the actions of living-beings [7]) and display of emotions and sociability (both supported by ethological studies [1,6,19]). We also added an item for personality, as it is a trait that has been identified in animals [11].

From a different perspective, Thomas and Johnston argue that cognitive processes can be illustrated through expressions [35]. They explain, in fact, that workers at Disney, throughout its early years, extensively studied animals, concluding they “communicate their feelings with their whole body attitude and movement” [35]. This was the origin of the 12 Principles of Animation, 9 of which were incorporated into our scale. Particularly, we did not consider those detailing how to make appealing characters, 2D animations and narratives as these were beyond the scope of our assessment.

In the case of biology, living beings are viewed as open systems [14]: they retrieve matter and energy from the environment and, in return, perform actions and produce waste material. From this perspective, a living entity is expected to, at the very least, grow in size and number of cells, and reproduce sexually or asexually to generate offspring.

Within the scope of illustration, Whitlatch identifies the importance of internal and external coherence [40,41]: the coherence between a creature’s behavior and their body’s design (“why a creature looks the way it does” [41]) and, likewise, between its body and its habitat (“the anatomical structure supports and makes possible the lifestyle” [41]).

The research developed up to this point resulted in a list of 46 statements (revised once for form simplification) depicted in Table 2.

2.2 Administering the Initial Creature Believability Questionnaire

To administer the initial iteration of the scale, we designed a questionnaire¹ where participants, after filling out demographic data, are shown 28, 20 to 30 seconds long, clips, obtained from various videogame sources. These clips all included at least one creature engaged in a specific activity. We used short clips because we meant to incite immediate responses rather than evaluating the participant's recollection of the clips (which could be contaminated by memory inaccuracy).

After viewing each clip, subjects were prompted to:

1. Describe how many creatures were present in the video. This was included as a fail-safe measure [15, 20] to allow deployment to Mechanical Turk.
2. Score 6 statements, taken from our believability items, through a 5-point Likert Scale. These items were displayed in a randomized order. Moreover, to reduce confirmation bias, 3 of the 6 items refer to elements absent in the respective clip. There were two reasons behind showing only a subset of the scale. The first involves user fatigue as administering the complete questionnaire would account for 1288 items (46×28). While one could argue that increasing the video length would potentially reduce the number of clips and subsequently the items' total, it would force the users to recollect more information, something we previously stated we wanted to avoid. The other reason was due to the fact that, to our knowledge, not every game has creatures with the characteristics present in the scale's items. However, grouping creatures from different games together would, on one hand, cause the issue we previously discussed and, on the other, potentially introduce a bias due to the change of context between games.
3. Rate the clip's creatures' believability using a 10-point Differential Semantic Scale with a Non-Believable-Believable pair. This was meant to assess the presence of correlations between Likert items and the creature's believability.
4. Similarly rate the clip's setting's believability. This was to reveal whether or not (and to what extent) the setting's and the creature's perceived believability are correlated.

2.3 Choosing Content for Evaluation

The games, included in this experiment, were chosen, by taking into account the following factors: first, we selected ones where creatures had an extensive on-screen presence, as we assume that, in these games, creatures have an additional development effort that would not be justified otherwise. As such, most of the games we considered are ones with open-world elements and life-simulation games. Finally, we chose to consider games made in the last 15 years. The main reason behind this lies in our belief that such games incorporate recent technology. This way, we mean to reduce any bias which could arise from notorious technological limitations.

This selection process resulted in several creatures from 19 games:

¹ <https://goo.gl/forms/XNe60psywQIILq302>.

- Hyenas and cheetahs from Afrika [26]
- D-Horse and D-Dog from Metal Gear Solid V: The Phantom Pain [16]
- The EyePet from EyePet [29]
- The dog from Fable 2 [17]
- Dogmeat from Fallout 4 [4]
- A rhino from Far Cry 4 [37]
- An Adamantoise from Final Fantasy XIII [33]
- Chop from Grand Theft Auto V [27]
- A black panther and a Bengal tiger from Kinectimals [9]
- A Rathian and a Rathalos from Monster Hunter Freedom Unite [5]
- The Artic Fox from Never Alone [38]
- A dog from Nintendogs [24]
- Red Pikmins from Pikmin [23]
- Dogs and a cat from The Sims 2 [21]
- A sabertooth tiger from The Elder Scrolls V: Skyrim [3]
- Chaos from Sonic Adventure 2: Battle [30] and Sonic Adventure DX: Director’s Cut [31]
- Creatures from Spore [22]
- Trico from The Last Guardian [10]

Before deploying to a larger population, the survey then underwent a pilot testing process with 5 test subjects, correcting typos and other errors.

3 Results and Scale Revision

The following subsections detail the process used in the revision of the believability scale.

3.1 Process and Population Profile

Our survey was deployed to Mechanical Turk, where 43 users participated (32 Males and 11 Females) with an average age of 31 ± 6 . Regarding education, 19% of the participants had an Highschool degree, 70% had a Bachelor’s degree whilst 12% had a Master’s degree. Finally, 35% of these users had a weekly exposition to media (videogames, movies, tv) of up to 20 h, while others had 20 to 40 weekly hours (42%), 40 to 60 weekly hours (16%), or 60 to 80 weekly hours (7%).

The results’ analysis was performed under a two-step method. Firstly, we analyzed the items on a per-clip basis. This allowed us to study how the believability scores could correlate with the clip’s items and remove the ones who did not. This is explained in Subsect. 3.2. Secondly, we grouped the items together and performed a reliability and factor analysis, on the questionnaire as a whole, as depicted in Subsect. 3.3.

3.2 Analysis 1

Before analyzing the items on a per-clip basis, we first studied the reliability of the believability semantic differential scales, which will be henceforth named as believability ratings. Specifically, we grouped them together and then calculated their Cronbach Alpha coefficient. As expected, results show a value of 0.96. While this is considered redundant [34], it is as predicted since the group consisted in the same question across all clips.

Having an indication that the believability ratings were internally consistent, we ran a Principal Components Analysis (PCA) on each group of items (the 3 non-control items plus its corresponding believability rating). By fixing one factor, we considered to be Believability, we used the believability ratings as a control value: if the result loaded in factor as well as other items, they were assumed to measure the same construct and thus, were kept for the next analysis. Furthermore, we used cut-threshold loading value of 0.4. The resulting items are indicated in the column “Analysis 1” of Table 2.

As depicted in the table, most items were able to load alongside the believability ratings. In fact, only 6 items were left out because their loadings were inferior to 0.4. These were item 16 (The creatures have diverse priorities), item 21 (The creatures feel empathy towards other creatures), item 26 (The creatures work with other creatures for a common goal), item 32 (The creatures absorb substances/energy from the environment to survive), item 40 (The creatures have traits particular to their sex) and item 46 (The creatures play with others).

3.3 Analysis 2

Having now a filtered list of items, we grouped the remainder 40 items together and analyzed them as a whole questionnaire. First, we analyzed the group’s internal consistency by calculating its Cronbach Alpha coefficient. With a value of 0.9, between the accepted values [34], we did not, at this point, remove any further item.

We then performed an exploratory factor analysis. The technique used was PCA with a Varimax Rotation, utilizing Eigenvalues ≥ 1 as a stopping criteria. This resulted in 11 components, illustrated in Table 1. However, the number as being too many for practical application of the scale. In order to find a more satisfactory solution, one with less factors, we established an additional criteria: a total variance explained minimum value of around %50. As expressed in Table 1, this accounts for retaining either 4 or 5 components.

Our next step involved choosing between 4 or 5 factors. To this extent, we ran two additional PCA, with the same rotation method as before, one by fixing 4 and the other 5 factors. However, this time, we used a cut threshold loading factor value of 0.4. From observing the resulting loaded items, we concluded that using 4 factors, over 5, grouped items with similar underlying semantics and thus, would facilitate the process of naming/categorizing those factors. The factor loadings, resulting from a PCA with 4 fixed factors, are described in Table 2.

Table 1. The extracted components using a PCA, with a Varimax Rotation and a stop criteria of Eigenvalue ≥ 1 .

Component	Eigenvalue	% of Variance	Cumulative %
1	9.067	23.859	23.859
2	3.551	9.346	33.205
3	3.118	8.205	41.410
4	2.765	7.276	48.686
5	1.923	5.061	53.747
6	1.875	4.933	58.680
7	1.693	4.455	63.135
8	1.471	3.870	67.005
8	1.393	3.667	70.672
10	1.306	3.437	74.109
11	1.119	2.945	77.054

Next, three of the researchers analyzed independently the factors’ loadings in order to find categories to represent each factor. This process underwent as follows:

1. Each of the three researchers individually studied the obtained factors and corresponding loadings and came up with naming proposals which would explain most, if not all, of the loaded items. This included deciding in which factors cross-loading variables would be kept.
2. We then gathered to discuss our proposals. During this step, we considered discarding items which deviated from our proposed semantics.
3. The process ended when we reached a consensus.

This processed originated the Relation with the Environment, Biological/Social Plausibility and Sociability, Adaptation and Expression concepts for explaining factors 1 through 4 respectively. From this process, besides removing items with factor loadings below 0.4, an additional 5 items were removed. This included item 6 (The creatures’ actions involve more than one step) and 17 (The creatures alternate between tasks) because their underlying concept did not align with the other factor-adjacent items; items 19 (The creature show expressions to known stimuli) and 41 (the creatures’ postures and expressions are coherent with their behavior) who appeared to be better suited for loading with the Expressions factor; and, finally, item 44 (the creatures’ bodies and behaviors are consistent) which we considered to belong to the Biological/Social Plausibility factor. The final scale, and encompassing items, are then as follows:

1. **Relation with the Environment** - This category corresponds to the items related to environment interactions, ranging from reactions to environmental cues or directed behaviors to systemic exchanges. The originated items are then as follows:

Table 2. The Believability Scale Analysis. The first column depicts the original scale’s items after their phrasal revision. The second column display which items were kept during the per-clip PCA with one fixed factor and a loading cut-threshold value of 0.4. The third column shows the loadings obtained after the second analysis performed using a PCA, with 4 fixed factors, Varimax Rotation and a loading cut-threshold value of 0.4. Values are omitted when their loading fails to meet the threshold. The items kept on the final iteration of the scale are depicted in bold.

	Item	Analysis 1 (Passed)	Analysis 2			
			1	2	3	4
1.	The creatures move by themselves	✓	0.442	0.513		
2.	The creatures’ motions are fluid	✓			0.516	
3.	The creatures’ motions reflect their weight/size	✓		0.464		
4.	The creatures’ expressions anticipate their actions	✓				0.539
5.	The creatures make several simultaneous motions	✓		0.714		
6.	The creatures’ actions involve more than one step	✓	0.622			
7.	Each of the creatures’ body parts have inertia	✓				
8.	The creatures interact with the environment	✓	0.617			
9.	The creatures react to stimuli	✓		0.735		
10.	The creatures controls their body	✓	0.587			
11.	The creatures recognize themselves	✓				
12.	The creatures’ behaviors differ from other creatures of the same species	×				
13.	The creatures focus on stimuli	✓		0.620		
14.	The creatures direct their behaviors towards targets	✓	0.763			
15.	The creatures locate objects in the environment	✓	0.514	0.454		
16.	The creatures have diverse priorities	×				
17.	The creatures alternate between tasks	✓			0.580	
18.	The creatures show positive (or negative) emotions towards objects, or events	✓				0.593
19.	The creatures show expressions to known stimuli	✓	0.416			0.691
20.	The creatures express moods through their body	✓				
21.	The creatures feel empathy towards other creatures	×				
22.	The creatures’ same-stimuli reactions change over time	✓			0.668	
23.	The creatures learn from past events	✓			0.544	
24.	The creatures learn through imitation	✓				0.620
25.	The creatures are able to apply old behaviors to new, similar, situations	✓			0.444	0.564
26.	The creatures use objects from the environment	×				
27.	The creatures make tools	✓				
28.	The creatures work with other creatures for a common goal	×				
29.	The creatures coordinate with other creatures	✓		0.744		
30.	The creatures communicate with other, same-species, creatures	✓		0.723		

(continued)

Table 2. (*continued*)

31.	The creatures communicate with other, different-species, creatures	✓			0.519	
32.	The creatures absorb substances/energy from the environment to survive	×				
33.	The creatures expel material	✓	0.632	0.449		
34.	The creatures have different life-stages	✓				
35.	The creatures change the way they look with age	✓			0.682	-0.425
36.	The creatures change the way they sound with age	✓			0.583	
37.	The creatures change the way they behave with age	✓			0.724	
38.	The creatures engage in reproductive acts	✓		0.539		
39.	There are signs of previous reproductive acts, such as eggs, cubs, pregnancy, etc	✓		0.495	0.433	
40.	The creatures have traits particular to their sex	×				
41.	The creatures' postures and expressions are coherent with their behavior	✓	0.621			
42.	The creatures' actions are appropriate to their context	✓	0.654			
43.	The creatures' body are adapted to their habitat	✓				0.694
44.	The creatures' bodies and behaviors are consistent	✓	0.702			
45.	The creatures play by themselves	✓				
46.	The creatures play with others	×				

- The creatures interact with the environment
- The creatures controls their body
- The creatures direct their behaviors towards targets
- The creatures locate objects in the environment
- The creatures expel material
- The creatures' actions are appropriate to their context

2. **Biological/Social Plausibility** - Corresponds to the creature's plausibility as a biological organism. This is demonstrated by showing autonomy and reactivity to its surroundings. Additionally, this category also encompasses the creature's ability to interact with other creatures. The items are as follows:

- The creatures move by themselves
- The creatures' motions reflect their weight/size
- The creatures make several simultaneous motions

- The creatures react to stimuli
 - The creatures focus on stimuli
 - The creatures coordinate with other creatures
 - The creatures communicate with other, same-species, creatures
 - The creatures engage in reproductive acts
 - There are signs of previous reproductive acts, such as eggs, cubs, pregnancy, etc.
3. **Adaptation** - This category involves learning behaviors and growing (which we considered to be an adaptation at the biological level). The originated items are as follows:
- The creatures' same-stimuli reactions change over time
 - The creatures learn from past events
 - The creatures are able to apply old behaviors to new, similar, situations
 - The creatures change the way they look with age
 - The creatures change the way they sound with age
 - The creatures change the way they behave with age
4. **Expression** - Expression encompasses the elements wherein creatures use their body as a means to communicate, learn or survive. The items are as follows:
- The creatures' expressions anticipate their actions
 - The creatures show positive (or negative) emotions towards objects, or events
 - The creatures show expressions to known stimuli
 - The creatures learn through imitation
 - The creatures' body are adapted to their habitat

Finally, we performed an additional reliability test on the remainder items as a confirmation. By calculating the Cronbach Alpha coefficient, it yielded 0.88 which is inside the acceptable range [34].

4 Conclusion

In sum, we presented the initial design and validation of a Believability Assessment scale, meant to be applied to videogame creatures. This included the process underlying the scale's items' generation as well as the validation of the scale as a whole.

We administered the scale as a questionnaire involving the visualization of videogame-related short clips. The answers were then analyzed using a two-step process. Firstly, the answers were analyzed on a per-clip basis, in order to study how the believability scores could correlate with the clip's items and filter them accordingly. Secondly, the remainder items were grouped together so reliability and factor analysis could be performed. Finally, these results were used to revise the scale. After revision, out of the original 46 devised items, 26 were kept, divided among 4 factors.

However, this scale still needs further research. While the exploratory factory analysis suggested how the scale could be divide among several dimensions, the

resulting structure still needs validation. Thus, Confirmatory Factor Analysis is still required as future work. Once the structure has been validated, a spectrum can be constructed out of an ordered set of creatures, from several videogames, to be used as a case study.

The creation of such a scale provides an insight into creature design, as the existing literature on believability is either focused on humans, narratives, or follows two distinct directions: one centering on behaviors whilst the other revolving around expressions. With our scale, we attempt not only to provide a tool to assess believability on non-human creature, but also one which unifies several perspectives on how to convey believability.

Acknowledgements. This work was supported by the FCT Ph.D. Grant SFRH/BD/100080/2014.

References

1. Adkins-Regan, E.: *Hormones and Animal Social Behavior*. Princeton University Press, Princeton (2005)
2. Bates, J.: The role of emotion in believable agents. *Commun. ACM* **37**, 122–125 (1994)
3. Bethesda Game Studios: *The Elder Scrolls V: Skyrim*. [DVD-ROM] (2011)
4. Bethesda Game Studios: *Fallout 4*. [DVD-ROM] (2015)
5. Capcom: *Monster Hunter Freedom Unite*. [UMD] (2008)
6. Darwin, C.: *The Expression of the Emotions in Man and Animals*. John Murray, London (1889)
7. Dennett, D.: Intentional systems theory. In: *The Oxford Handbook of Philosophy of Mind*, pp. 1–22. Oxford University Press (2009)
8. Fogg, B., Tseng, H.: The elements of computer credibility. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (1999)
9. Frontier Developments: *Kinectimals*. [DVD-ROM] (2010)
10. genDESIGN and SIE Japan Studio: *The Last Guardian*. [Blue-Ray] (2016)
11. Gosling, S.D.: Personality in non-human animals. *Soc. Pers. Psychol. Compass* **2**, 985–1001 (2008)
12. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. *MIS Q.* **28**, 75–105 (2004)
13. Hingston, P.: *Believable Bots: Can Computers Play Like People?*. Springer, Heidelberg (2011)
14. Kadhil, N.: Characteristics of living organisms. In: *Biology*, pp. 1–10. Cambridge University Press (2007)
15. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical turk. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2008*, pp. 453–456. ACM, New York (2008)
16. Kojima Productions: *Metal Gear Solid V: The Phantom Pain*. [Blu-Ray] (2015)
17. Lionhead Studios: *Fable II*. [DVD-ROM] (2008)
18. Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. *J. Comput. Mediated Commun.* **3**(2), 0–39 (1997)
19. Lorenz, K.Z.: *The Foundations of Ethology*. Springer, New York (2010)
20. Mason, W., Suri, S.: Conducting behavioral research on amazon’s mechanical turk. *Behav. Res. Methods* **44**(1), 1–23 (2012)

21. Maxis: *The Sims 2*. [DVD-ROM] (2005)
22. Maxis: *Spore*. [DVD-ROM] (2008)
23. Nintendo EAD: *Pikmin*. [Mini Disc] (2001)
24. Nintendo EAD: *Nintendogs*. [Flash Card] (2005)
25. Raúl, A., Ledezma, A., Sanchis, A.: ConsScale: A pragmatic scale for measuring the level of consciousness in artificial agents. *J. Conscious. Stud.* **17**, 131–164 (2010)
26. Rhino Studios: *Afrika*. [Blu-Ray] (2008)
27. Rockstar North: *Grand Theft Auto V*. [Blu-Ray] (2013)
28. Rosenkind, M., Winstanley, G., Blake, A.: Adapting bottom-up, emergent behaviour for character-based AI in games. In: Bramer, M., Petridis, M. (eds.) *Research and Development in Intelligent Systems XXIX*, pp. 333–346. Springer, Heidelberg (2012)
29. SCE London Studio and Playlogic Game Factory: *Eye Pet*. [Blu-Ray] (2009)
30. Sonic Team USA: *Sonic Adventure 2: Battle*. [Mini Disc] (2002)
31. Sonic Team USA: *Sonic Adventure DX: Director's Cut*. [Mini Disc] (2003)
32. Spector, P.E.: *Summated Rating Scale Construction: An Introduction*. Sage Publications, Thousand Oaks (1992)
33. Square Enix 1st Production Department: *Final Fantasy XIII*. [Blu-Ray] (2010)
34. Steiner, D.L.: Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* **80**(1), 99–103 (2003)
35. Thomas, F., Johnston, O.: *The Illusion of Life: Disney Animation*. Hyperion (1997)
36. Togelius, J., Yannakakis, G.N., Karakovskiy, S.: Assessing believability. In: *Believable Bots: Can Computers Play Like People?*, pp. 215–230. Springer, Heidelberg (2011)
37. Ubisoft Montreal: *Far Cry 4*. [DVD-ROM] (2014)
38. Upper One Games: *Never Alone*. [Digital] (2014)
39. Warpefelt, H.: *The Non-Player Character Exploring the believability of NPC presentation and behavior*. Ph.D. thesis, Stockholm University, Faculty of Social Sciences, Department of Computer and Systems Sciences, Stockholm, Sweden (2016)
40. Whitlatch, T.: *Principles of Creature Design: Creating Imaginary Animals*. Design Studio Press, Culver City (2015)
41. Whitlatch, T.: *Science of Creature Design: Understanding Animal Anatomy*. Design Studio Press (2015)
42. Wobbrock, J.O., Kientz, J.A.: Research contributions in human-computer interaction. *Mag. Interact.* **23**(3), 38–44 (2016)