

In Search of the Uncanny Valley

Frank E. Pollick

Department of Psychology, University of Glasgow,
Glasgow, UK

frank@psy.gla.ac.uk

Abstract. Recent advances in computer animation and robotics have led to greater and greater realism of human appearance to be obtained both on screen and in physical devices. A particular issue that has arisen in this pursuit is whether increases in realism necessarily lead to increases in acceptance. The concept of the uncanny valley suggests that high, though not perfect, levels of realism will result in poor acceptance. We review this concept and its psychological basis.

Keywords: uncanny valley, animation, robot.

1 Introduction

This essay reviews the concept of the uncanny valley which clearly states that increased realism does not necessarily imply acceptance. This review is not alone and related writings on the uncanny valley can be found elsewhere [1-4], as well as discussion of its prominent role in considerations of robots which are human-like in appearance [1, 5-7]. Thus, any review runs the risk of repeating what has already been said. To attempt to avoid this I have tried to focus on the psychology of the uncanny valley and to discuss what psychological principles might underlie its existence. With this in place we can look at falling into the uncanny valley not from the usual perspective of ever more realistic artifacts, but instead from the viewpoint of how normal human activity might be modulated to fall into the same uncanny valley.

The body of this essay is contained in four parts: The first three discuss, in turn, the history of the uncanny valley, evidence for its existence and theoretical arguments for its plausibility. The final section provides an operational definition of the uncanny valley that is examined in the context of human behavior, and the shortcomings which arise are discussed.

2 History

In 1970 Dr. Masahiro Mori, a Professor of Engineering at Tokyo Institute of Technology, put forth the following thought experiment: Assume we could make a robot more and more similar to a human in form, would our affinity to this robot steadily increase as realism increased or would there be dips in the relationship between affinity and realism [8]. Mori put forth the proposition that the latter would be the case - as the robot became more human-like there would first be an increase in its acceptability

and then as it approached a nearly human state there would be a dramatic decrease in acceptance. He termed this precipitous drop “bukimi no tani” and the translation of “bukimi no tani” into “uncanny valley” has become popularized. The hypothesized shape of the uncanny valley revealed in the relationship between affinity and realism is shown in Figure 1.

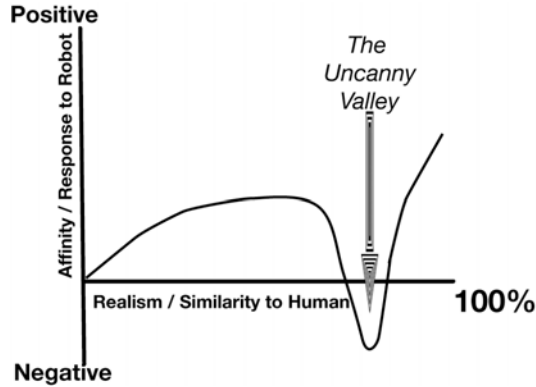


Fig. 1. Simplified diagram showing the hypothesized relationship between affinity and realism with the uncanny valley appearing as a negative response as one approaches total realism

Mori also appreciated that robots are not defined by the single dimension of form and considered the effect of robot motion. Here he proposed that motion and form together will produce a different function of acceptability versus realism. For example, motion could deepen the valley since form sets up expectations in an observer and if other factors such as motion do not match these expectations then there is further rejection of the entity. Mori tread a little further into the realm of thought experimentation and illustrated this phenomenon with the example of viewing a corpse. Certainly a corpse has to be very similar in form to an actual human yet we find viewing a corpse unpleasant, and if the corpse would suddenly move to stand up we would be terrified. Motion could also be used to circumvent a fall into the uncanny valley as Mori illustrated with the example of Japanese bunraku puppets. These puppets are somewhat basic in form and are accompanied by their black-cloaked puppeteers as they appear on stage, however their lifelike motion leads them to be accepted as nearly human as one follows the action on stage.

The practical significance, and lingering influence of Mori’s proposal of the uncanny valley is found in his suggestion to designers of robots and other related artifacts. He proposed that the first peak in acceptability is an effective target for design. Here there are moderately high values of acceptance and a safe distance from the uncanny valley. Striving for realism will only lead to the risk of catastrophic tumbling into the uncanny valley. The imaginative example of zombies is transformed into advice for designers of artificial limbs that if they cannot get the motion, texture and temperature of an artificial hand to be correct then having it look perfect could lead to awkward situations when the artificiality is discovered in a social interaction like a handshake. A characteristic of Mori’s advice to designers is that he did not

provide precise definitions of realism or affinity and thus the concept of an uncanny valley has proven to be a broadly applicable guidepost to designers in a variety of domains.

Before leaving a discussion of historical aspects of the uncanny valley it is worth briefly considering the term “uncanny valley” itself and its translation from the Japanese “bukimi no tani”. Here “tani” is quite directly “valley” and “no” is a connecting particle, while “bukimi” has several translations including “eery”, “strange” and “uncanny”. How “uncanny” was chosen for “bukimi” is an interesting question. The first appearance of this translation appears to be in the book “Robots: Fact, Fiction and Prediction” [9]. We can speculate that “uncanny” was chosen due to its psychological resonances with the 1919 essay of Sigmund Freud entitled “Das Unheimlich” which was translated to “The Uncanny” [10]. While “unheimlich” appears to be one of those problematic words which defy a simple translation, Freud specified the uncanny as that class of the frightening, which leads back to what is known of old and long familiar. This sense of the word seems particularly appropriate to the phenomenon Mori described as the valley appears as we approach the familiar.

3 Evidence for Existence

The original writing of Mori in Japanese referring to “bukimi no tani” and its translation into English as the “uncanny valley” both occurred during the 1970s. However, the concept seems to have laid dormant for almost 30 years and resurfaced as technology inched towards increasing levels of sophistication in computer graphics and robotics. This increase in sophistication makes it possible for greater and greater realism to be attained. However, increased realism has not necessarily equated with increased acceptance by the public and the existence of an uncanny valley has been called on to describe this phenomenon.

The uncanny valley entered the popular lexicon not long after the full-length feature film *Final Fantasy* appeared. This film consisted entirely of characters generated by computer graphics and used high levels of realism. The audience response was lukewarm and a general consensus began to evolve that it failed due to falling into the uncanny valley. Andy Jones, the *Final Fantasy* animation director gives a telling quote in *Wired* magazine when he says “it can get eerie. As you push further and further, it begins to get grotesque. You start to feel like you’re puppeteering a corpse” [10, 11]. Film critic Roger Ebert continues this thread in his column in the *Chicago Times* when discussing the role of Andy Serkis in portraying the character Gollum in the *Lord of the Rings: Return of the King*, he says “If Serkis brought Gollum to life, other artists fine-tuned the balance with the uncanny valley” [12]. A final discussion point is a comparison of the computer-animated films *Polar Express* and *The Incredibles* which both opened in fall 2004. *Polar Express* featured realistic animation while *The Incredibles* was more stylistic. The initial reluctance of audiences to view the *Polar Express* began to be accounted for by its tumble into the uncanny valley while *The Incredibles* avoided this fate [13]. These examples are compelling in suggesting that the uncanny valley exists, however, they still do little to explain what the uncanny valley is and what conditions are critical for its occurrence.

Recently the field of robotics has moved towards robots designed to leave the assembly room floor and work alongside humans [14-16]. The most ambitious of such designs are humanoid robots that are modeled upon human structure and androids that strive for greater similarity to human form and function. Justifications for mimicking human form include that teaching the robot by demonstration might be facilitated by the teacher and robot having the same structure. Additionally, the robot will be able to function in spaces designed for humans and with human tools and that this would eliminate the need for special design considerations for the robots. A survey of such humanoid robots as the Honda Asimo, Sony Qrio and Toyota's partner robot reveal that all present a distinctly artificial appearance. At least for Qrio, this appearance is intentional as revealed in an interview with Toshitada Doi about Qrio presented on the Sony web pages. When asked – "What do you think about the "character" of robots? – he answered "Take QRIO as an example. We suggested the idea of an "eight year-old space life form" to the designer -- we didn't want to make it too similar to a human. In the background, as well, lay an idea passed down from the man whose work forms the foundation of the Japanese robot industry, Masahiro Mori: "The Valley of Eeriness". If your design is too close to human form, at a certain point it becomes just too . . . uncanny. So, while we created QRIO in a human image, we also wanted to give it little bit of a "spaceman" feel." [17].

While much of the evidence to support the uncanny valley, like that above, is anecdotal there have been limited attempts to experimentally confirm its existence. The primary evidence to support its existence comes from research by MacDorman and Ishiguro [1] that explored observers reactions to facial morphs from a mechanistic robot - to a human looking robot – to an actual human. What they found was that at the boundary of the mechanistic robot and the human looking robot there was a rise in judgments of the eeriness of the display that was consistent with judgments of the morph being seen as less human. However, using the same technique of morphing and identical stimuli as the bases for the morphing space, David Hanson has asked the question of whether falling into the uncanny valley is inevitable [18]. What he did first was to replicate the findings of MacDorman and Ishiguro to find a peak in eeriness judgments. What he did next was to "tune" the different morphs so that they would appear more attractive. What he found was that the eeriness ratings were a flat line although there still was a distinct transition of ratings from human to nonhuman. This clearly indicates that for the case of the single cue of appearance, uncanny reactions can be circumvented by skillful manipulation. It is possible that such a process can be extended to multiple cues if their complex interactions did not make the tuning process intractable. Another experiment investigating the basis of the uncanny valley has shown that the inanimate features of human-like robots which denote death could instill responses consistent with a fear of death in observers [1]. Finally, results from Ishiguro [19] have shown that an android robot undergoing small movements equivalent to postural adjustments could be viewed for 2 seconds without an observer detecting that they were viewing an artificial agent. Without motion observers were much more likely to detect that the agent was not human. This indicates that for very brief encounters the uncanny valley can be avoided without difficulty.

The preceding discussion brings into focus the current situation regarding information about the existence of the uncanny valley. It can be seen that the

examples from feature films indicate that the uncanny valley exists. Moreover, many robot designers, animators and game designers appear sufficiently respectful of the concept that they design away from the uncanny valley. However, it can be argued regarding feature film that the uncanny valley is being used as a catch-all phrase when a realistic animation fails. Moreover, the limited empirical evidence both restricts extensive conclusions being drawn and further suggest that falling into the uncanny valley is not inevitable.

4 Psychological Plausibility

If the uncanny valley exists then it should be possible to explain why it exists and possibly to mitigate its effects. Such an explanation doesn't yet fully exist, however we can examine various proposals and related research findings to estimate the plausibility of the phenomenon and explanation. There are at least four descriptions of relevant psychological processes that could predict the uncanny valley and they will be discussed in the following paragraphs.

One common explanation is related to the perceptual issue that increased realism seems inextricably linked to increased information and thus if there are errors in our approximations to realism then they might simply become more evident as more information is provided. One issue with this explanation is that it begs the question as to *why* the errors would become more apparent. It would be just as easy to predict that with the greater and greater amounts of generally correct information being presented that any incorrect information would be drowned out. If this doesn't happen then there must be a peculiar sensitivity to the information which is incorrect. Thus, while this explanation has a ring of truth, it does not appear to be a complete explanation.

A cognitive issue noted by Ramey [20] is that although the uncanny valley is modeled to lie along a continuum of realism, the appreciation of what is being viewed lies at a categorical boundary between humans and machines. Since processes of event and object *categorization* are obligatory, the uncanny valley then is predicted once a lack of genuineness is discovered, the clever animation or robot seems not to fit solidly into either the living or non-living category. This inability to categorize will then lead to a state of dissonance. It appears that this cognitive issue of classification cannot be avoided, however since category boundaries are not necessarily static, the possibility then arises that increasing exposure will lead to a third category being developed which resolves the dilemma.

Another possible explanation for how the uncanny valley might come about is a refinement of the first proposal and inspired by the observation of Mori that motion could exacerbate an uncanny situation already existing in form. The generalization of this idea is that human actions consist of a wealth of different sensory cues. If these cues are not mutually consistent then reconciling the differences among cues might lead to a state of unease or at least uncertainty about what is being observed. The case of form and motion are interesting since various research leads to the view, consistent with the observation of Mori, that there are separate visual pathways that initially process form and motion information and then at a later stage integrate this information in the process of representing human actions [21]. One implication of

this is that form and motion might contain different cues to human activity, a view supported by experimental results which indicate that the recognition of affect and emotion from human movement is represented by dimensions of activity and valence (positivity/negativity). The dimension of activity is supported by the speed of a movement and the valence dimension appears to be related to structural relations among the body parts [22]. These arguments point towards the question of whether the uncanny valley could arise out of mismatches between sensory cues where the subtle inconsistencies between cues or missing inputs might lead to finding an experience unpleasant. Certainly, one testable claim about Mori's presentation of the uncanny valley would be to find an uncanny form and to see if motion can be used to modulate the experience.

The previous paragraphs took the position that human activity forms a multidimensional signal and that an uncanny valley might come about in artificial systems either due to a subtle disorganization of the information carried along these dimensions or the subsequent difficulties on categorizing an event that falls on a category boundary. These explanations are not mutually exclusive but the emphasis on the available information and its categorization leaves out one potentially important aspect that has gained increasing interest in the field of neuroscience. This is that our sensitivity to particular information and its subsequent classification is driven by social (and survival) needs to communicate and react to the individuals around us. Related research is asking the parallel question of what brain processes are involved during observation of another social agent (another human) or a non-social agent (a robot). At present the results are inconclusive, some researchers find different responses to humans versus robots at the brain (Tai *et al.*, 2004) as well as behavioral levels [23, 24]. Other results find that both robot and human movements elicit automatic imitation [25], and finally some find mixed results in comparing responses to human versus artificial agents [26]. It is early days and this research has yet to delve into the uncanny valley but it is asking a key question regarding how a "social" brain evaluates its environment. That is to say that the critical issue might not be the logical problem of evaluating human versus nonhuman or confusion over a mismatch of perceptual cues. Rather, the issue might be how the social brain evaluates these perceptual cues and cognitive scenario. Support for this view can be seen in studies showing differences in the acceptance of robots by different cultures [27] and across the lifespan [28].

The purpose of this section was to review principles from psychology that might lead to the prediction of an uncanny valley. Several concepts were presented which suggest that from the standpoint of psychological theory it is plausible that an uncanny valley would exist. However, an explanation of the uncanny valley did not appear to be the providence of any one unique concept.

5 The Human Side of the Uncanny Valley

Thus far I have discussed the original thought experiment of Mori that introduced the uncanny valley, described evidence for its existence including widespread acceptance as a design principle in robotics and computer animation, and put forth psychological concepts that argue for its plausibility. This suggests a definition of the uncanny

valley as a phenomenon that exists in the stimulus space around normal human activity and is triggered from either perceptual mismatches or categorical effects, but that the critical level of evaluation might be social. This definition avoids an obvious and important question about how to precisely characterize the dimensions of realism and affinity used in the plots of the uncanny valley. However, this issue is possibly best addressed only through empirical investigations. What I want to examine now is the question of how exclusive is this definition. In particular, if instead of starting at a cute robot toy and moving towards the uncanny valley, what if we start with human activity and move towards the uncanny valley. To do this I will briefly examine three phenomena which seem to fit different criteria of the proposed definition. These phenomena include dubbed speech in cinema, fear of clowns and Capgras syndrome.

The first example of dubbed speech satisfies the property of being in the vicinity of normal human activity since it combines an actual human movement with an actual auditory signal from a different language which is not entirely congruent. This would lead to both perceptual mismatches as well as the possibility for categorical effects of which language is being spoken. Recent evidence has described the importance of audiovisual processes in understanding speech [29] and shown that even very young infants are skilled at appreciating audiovisual congruence [30]. So the question should be then why doesn't an uncanny valley exist for dubbed speech? Perhaps it does. Evidence from a 1988 survey of British television viewers revealed that of those 32% viewers who prefer subtitling to dubbing, 42% did so because they dislike dubbed programs [31]. Moreover, a study of young children shown a subtitled and dubbed version of the same program preferred the subtitled version even though they would not have had advanced skill in reading [32]. Possibly dubbed speech is an example of where habituation, particularly in media markets which make frequent use of dubbing, can overcome a natural tendency to find the experience unpleasant.

The next example of clowns, while somewhat lighthearted, still seems a useful case. It can also be used to make the serious point, that although more seems to be written about the uncanny valley there is about as much empirical evidence to support clown phobia as the uncanny valley. Consulting the limited published report [33], the internet and informal interview does however reveal that those who hate clowns are not alone. Clearly a clown is just a human with some facepaint and funny clothes so they satisfy the condition that they are close to a normal human stimuli. Moreover, any categorical issues should be resolved by the category itself of "clowns" that makes it clear what kind of agent is being encountered. Possible perceptual inconsistencies include that the facial expression painted on the face is not consistent with the actions. Perhaps this incongruence might be appreciated more on a social level, that the clown with a painted smile ought not to always look so happy for all its actions.

The final example of Capgras syndrome suggest that an uncanny situation could arise without any perceptual mismatches and for normal human activity. Capgras syndrome is a relatively rare condition where the sufferer believes that people, or in some instances things, have been replaced with duplicates [34]. These duplicates are rationally accepted to be identical in physical properties but the irrational belief is held that the "true" entity has been replaced with something else. Ellis and Lewis

[35] describe the recent situation of a man who after a car accident believed that his wife had died in the accident, and the woman he currently lived with (his wife) was a duplicate. Naturally, he found this situation to be uncomfortable. Some sufferers of Capgras syndrome have even claimed that the duplicate is a robot and these cases would seem to perfectly match the uncanny valley. Ellis and Lewis [35] argue that the syndrome arises from an intact system for overt recognition coupled with a damaged system for covert recognition that leads to conflict over an individual being identifiable but not familiar in any emotional sense. This example provides support for a view that the uncanny valley could arise from issues of categorical perception that are particular to the specific way that the social brain processes information.

What this section has attempted to demonstrate is that there are sufficient possibilities for deviation from normal behavior and normal recognition to lead to scenarios consistent with a definition of the uncanny valley. Perhaps a more precise definition could avoid this multiplicity of ways into the uncanny valley, though this has the danger of throwing the proverbial baby out with the bathwater. Thus, we seem left with the situation that the obstacle to interpreting increased realism might not be one great uncanny valley but rather a multitude of uncanny potholes.

6 Conclusions

The goal of this essay was to review the uncanny valley and not to either refute or describe its essential mechanisms. The hope was that a description of its history, context and psychological plausibility would inform what questions are important to pursue. One essential question to ask is just whether there is enough evidence to say that the uncanny valley exists? Surprisingly, the answer is equivocal. It is clear from practitioners that more realism does not always equate with greater acceptance by audiences and there is a wealth of anecdotal evidence to support this view. Moreover, from first principles of psychology one can build a case that something like an uncanny valley would exist. However, there is a dearth of empirical evidence on the topic and certainly no study that outlines essential properties that can be manipulated to navigate into and out of the uncanny valley. Thus, it would seem some care is needed in the evaluation of claims about the uncanny valley until a more rigorous understanding is reached.

The attempt made here to come up with an operational definition of the uncanny valley ran into difficulties with its assumption that the uncanny valley should occur in the vicinity of natural human actions. Namely, this difficulty was that from the perspective of psychology there doesn't seem to be a shortage of situations where actual human actions can be transformed into the uncanny. What would be helpful to resolve this problem is further specification, by those animators and roboticists pushing into the uncanny valley, of what bit of the human response to these artifacts is the essential aspect of its uncanny nature. This does not seem a simple task since the uncanny region is at the cutting edge of technology and can be achieved only with substantial resources and talent and these are typically devoted to avoiding the uncanny valley.

References

1. MacDorman, K.F., Ishiguro, H.: The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies* 7(3) (2006)
2. Brenton, H., et al.: The Uncanny Valley: does it exist? In: The 11th International Conference on Human-Computer Interaction. Lawrence Erlbaum Associates, Las Vegas (2005)
3. Gee, F., Browne, W., Kawamura, K.: Uncanny Valley Revisited. In: 2005 IEEE International Workshop on Robots and Interactive Communication. IEEE Press, Nashville (2005)
4. Hanson, D., et al.: *Upending the Uncanny Valley* (2006)
5. Kosloff, S., Greenberg, J.: Android science by all means, but let's be canny about it! *Interaction Studies* 7(3) (2006)
6. Chaminade, T., Hodgins, J.: Artificial agents in the social cognitive sciences. *Interaction Studies* 7(3) (2006)
7. Canamero, L.: Did Garbo care about the uncanny valley? *Interaction Studies* 7(3) (2006)
8. Mori, M.: Bukimi No Tani (the Uncanny Valley). *Energy* 7, 33–35 (1970)
9. Reichardt, I.: *Robots: Fact, Fiction and Prediction*, p. 166. Thames and Hudson, London (1978)
10. Freud, S.: The Uncanny. The Standard Edition of the Complete Psychological Works of Sigmund Freud, vol. 17, pp. 219–252. The Hogarth Press, London (1960)
11. Weschler, L.: Why is this man smiling? Digital animators are closing in on the complex system that makes a face come alive. *Wired*, 2002(10.06)
12. Ebert, R.: Gollum stuck in 'Uncanny Valley' of the 'Rings', in *Chicago Sun Times*, Chicago (2004)
13. Horneman, J.: *The Incredibles, Polar Express, the Uncanny Valley*, Pixar (2004)
14. Atkeson, C., et al.: Using humanoid robots to study human behavior. *IEEE Intelligent Systems* 15, 46–56 (2000)
15. Coradeschi, S., et al.: Human-Inspired Robots. *IEEE Intelligent Systems* 21(4), 74–85 (2006)
16. Hale, J., Pollick, F.: Sticky Hands: learning and generalization for cooperative physical interactions with a humanoid robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 35(4), 512–521 (2005)
17. Sony, Interview with Toshitada Doi: Personal robots make the 21st century more fun (2006)
18. Hanson, D.: *Expanding the Aesthetic Possibilities for Humanoid Robots* (2005)
19. Ishiguro, H.: Android science: conscious and subconscious recognition *Connection Science* 18(3) (2006)
20. Ramey, C.: The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In: *Proceedings of the Views of the Uncanny Valley: workshop, IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan (2005)
21. Giese, M., Poggio, T.: Neural Mechanisms for the Recognition of Biological Movements. *Nature Neuroscience Review* 4, 179–192 (2003)
22. Pollick, F., et al.: Perceiving Affect from Arm Movement. *Cognition* 82, B51–B61 (2001)
23. Kilner, J.M., Paulignan, Y., Blakemore, S.J.: An interference effect of observed biological movement on action. *Current Biology* 13(6), 522–525 (2003)
24. Castiello, U.: Understanding Other People's Actions: Intention and Attention. *Journal of Experimental Psychology: Human Perception and Performance* 29(2), 416–430 (2003)

25. Press, C., et al.: Robotic movement elicits automatic imitation. *Cognitive Brain Research* 25(3), 632–640 (2005)
26. Pelphrey, K.A., et al.: Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience* 23(17), 6819–6825 (2003)
27. Kaplan, F.: Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics* 1(3), 465–480 (2004)
28. Turkle, S., et al.: Relational artifacts with children and elders: the complexities of cybercompanionship (2006)
29. Munhall, K., Vatikiotis-Bateson, E.: Spatial and Temporal Constraints on Audiovisual Speech Perception. In: Calvert, G.A., Spence, C., Stein, B.E. (eds.) *Handbook of Multisensory Processes*, pp. 177–188. MIT Press, Cambridge (2004)
30. Hollich, G., Newman, R., Jusczyk: Infants' use of synchronized visual information to separate streams of speech. *Child Development* 76(3), 598–613 (2005)
31. Kilborn, R.: 'Speak my language': current attitudes to television subtitling and dubbing. *Media Culture Society* 15(4), 641–660 (1993)
32. Koolstra, C.M., Peeters, A.L., Spinhof, H.: The Pros and Cons of Dubbing and Subtitling. *European Journal of Communication* 17(3), 325–354 (2002)
33. Austin, R., McCann, U.: Ballatrophobia: When clowns aren't funny. *Anxiety* 2, 305 (1996)
34. Ellis, H., et al.: Delusional misidentification: The three original papers on the Capgras, Frégoli and intermetamorphosis delusions. *History of Psychiatry* 5, 117–118 (1994)
35. Ellis, H., Lewis, M.: Capgras delusion: a window on face recognition. *Trends in Cognitive Science* 5(4), 149–156 (2001)