NEW FINDINGS SHED LIGHT ON A CENTURY’S WORTH OF BIZARRE EXPLANATIONS FOR THE EERIE FEELING WE GET AROUND LIFELIKE ROBOTS.

Into the Uncanny Valley
WHAT WE KNOW BY JOE KLOC / NOVEMBER 16, 2009

A dead body appears in almost every way to be a normal human. But the pallid skin and empty eyes signal that the person-shaped form we are looking at is, in a way we can’t even fully grasp, strange and disturbing.

We feel a similar eeriness when interacting with robots and models that look almost human but fall short of convincing us because of subtle peculiarities in their features. Poor box office returns on computer-animated films like “The Polar Express” and “Beowulf” were blamed on moviegoers finding the not quite true-to-life characters unsettling.

Disturbing experiences that feel both familiar and strange are instances of the “uncanny,” an intuitive concept, yet one that has defied simple explanation for more than a century. Interest in the particular occurrences of the uncanny, in which humans are bothered by interaction with human-like models, began as a psychological curiosity. But as our ability to design artificial life has increased—along with our dependence on it—getting to the heart of why people respond negatively to realistic models of themselves has taken on a new importance. Attempts to understand the origins of this reaction, known since the 1970s as the “uncanny valley response,” have drawn on everything from repressed fears of castration to an evolutionary mechanism for mate selection, but there has been little empirical evidence to assess the validity of these ideas.

New findings published in PNAS this September are putting some long-overdue experimental rigor behind the uncanny valley. Last spring at Princeton’s Neuroscience Institute, Asif Ghazanfar developed a computer model of a macaque monkey designed to interact with real macaques. But the monkeys weren’t fooled. Further testing revealed that, much to Ghazanfar’s surprise, his model was eliciting an uncanny valley response from the monkeys. It was the first time scientists had ever observed such a response in a non-human species.

“By showing that monkeys can do it, several things become plausible,” Ghazanfar says. “One is that there is an evolutionary explanation for the uncanny valley and the other is that it is not something specific to our human, cultural experience.” These finding may for the first time allow scientists to go back through a century’s worth of peculiar ideas about the origins of the uncanny valley and begin putting them to the test.

An Eerie History

Sigmund Freud offered the first major attempt to explain our uncanny response toward life-like human models. With World War I still dragging on across Europe in 1918, Freud was having trouble finding article submissions for his psychoanalytic journal, Imago, and so decided to contribute something himself. The following year, he published a bizarre 40-page essay on an almost completely unknown concept in psychology. Freud’s subject was the “uncanny,” a term coined 13 years earlier by a little-known German doctor named Ernst Jentsch.

Title “The Uncanny,” Freud’s essay is, in nearly every aspect, as strange as the phenomenon it struggles to understand. “There is a k of contradictory information in there,” says Samuel Weber, a professor of philosophy and literature at the European Graduate Schoo “If you put it together you realize it doesn’t add up neatly to any unified position.”

For instance, Freud begins with a disclaimer that he hasn’t had an uncanny personal experience in so long that he must “awaken in
[himself] the possibility of experiencing it,” implying that he either he wrote some 12,000 words about a psychological phenomenon he has no personal understanding of or he isn’t fully aware of his own familiarity with his subject. Weber sides with the latter interpretation. “It’s not a question of whether what he is doing is invalid, but whether there is more going on there than he wants to—or is able to—acknowledge,” he says.

Freud’s essay is, in nearly every aspect, as strange as the phenomenon it struggles to understand.

Indeed, Freud’s personal life often creeps into his examples of the uncanny. Such is the case as he explains that when he encounters a number—36 or 855, for instance—several times in the same day, he is overcome with an uncanny feeling. This is arguably one of the most universally shared uncanny experiences not involving an interaction with a human model. We are all intimately familiar with numbers, so when we encounter them in a strange context, we respond with a feeling of unease and suspicion.

What is telling about Freud’s use of this example is the number he chose to make his point: 62. According to his official biographer, Ernest Jones, Freud had written much of “The Uncanny” years before its publication but waited until he passed the age of 62 to complete it. Not coincidentally, Freud’s father was 62 when he died. Even while Freud denies any familiarity with the uncanny, he readily plucks examples from his personal life in order to illustrate it. The experience of writing “The Uncanny” must have been, for Freud, a rather uncanny one.

According to Freud, the phenomenon that would later be called the uncanny valley stems from a primitive attempt of humans to skir death and secure our own immortality by creating copies of ourselves—such as wax figures and, later, life-like robots. He quotes his colleague Otto Rank in saying that this “doubling” behavior is “an energetic denial of the power of death” and suggests the idea of the immortal soul was the first double of the body. Our uncanny response follows from the fact that most of us no longer believe we can secure our own immortality by making copies of ourselves, but we haven’t yet shaken the primitive habit of trying to do so. The sad consequence of this is that, in Freud’s words, “The double reverses its aspect. From having been an assurance of immortality, it becomes the uncanny harbinger of death.” The copies we feel compelled to make only serve to remind us why we began making them in the first place: We are, inevitably, going to die.

The Valley of Death

For Ghazanfar, Freud’s explanation of the uncanny valley, steeped in psychoanalytic theory, is much too “human-specific.” Nevertheless, the connection Freud makes between death and the uncanny valley persists in one form or another to this day.

For the most part, Freud’s essay reads like one big Freudian slip, revealing its author’s own anxieties about reconciling the uncanny with psychoanalysis. But in a sense, it succeeds despite itself: Its failures serve to illustrate the difficult nature of the uncanny, which is arguably the reason that for decades few scholars made serious attempts to investigate its origins: “It’s hard to treat the uncanny in the regular objectifying manner of the sciences or the humanities because it manifests itself through an interaction of subject and object—of feeling and situation—and in a way that is the hardest thing to analyze,” Weber explains.

In 1970 the Japanese roboticist Masahiro Mori published a short paper in the journal Energy in which he tried his hand at explaining the uncanny response we have toward human models. In much the same way Ghazanfar would later observe the uncanny valley response in monkeys, Mori noticed that when robots look very similar to us—but not so similar that we consciously mistake them for humans—our comfort level around them drops considerably. He dubbed this drop, bukimi no tani, or the “uncanny valley.”

In his paper, also titled “The Uncanny Valley,” he recommends that roboticists avoid building robots so realistic that they risk falling into the valley, offering the example of hands on a Buddha statue as an alternative approach to robot design: “The hand has no fingerprint, and it assumes the natural color of wood,” he wrote. “But we feel it is beautiful and there is no sense of the uncanny.”

“When we die, we fall into the trough of the uncanny valley.”
— Masahiro Mori

In the West, there is often a Frankensteinian stigma attached to artificial intelligence, but Mori offered Japan a much different perspective. In The Buddha in the Robot: A Robot Engineer’s Thoughts on Science and Religion, published in 1974, he wrote, “I believe robots have the Buddha-nature within them—that is, the potential for attaining Buddhahood.” His ideas about religion and the uncanny valley have had a substantial influence on the development of Japanese robotics. “In Japan, there is a great sensitivity in the government for having people who are accepting of robotics and robots in general. Mori’s interpretation of the uncanny valley became a kind of dogma,” says Karl MacDorman, a roboticist at Indiana University. As a result, Japan spent the next few decades avoiding human-like robot designs.
While the purpose of Mori’s paper was to inform robot design, in a concluding paragraph he cannot resist offering his own theory about the origins of the uncanny valley. He writes: “When we die, we fall into the trough of the uncanny valley. Our body becomes cold, our color changes, and movement ceases.” Human models fall into the uncanny valley because they remind us of death. “It may be important to our self-preservation,” he concludes.

Mori, like Freud, linked the uncanny valley to a “human-specific” notion of death, and many have suggested that he had Freud in mind when he penned “The Uncanny Valley”—which is possible since Freud’s concept of the uncanny, unheimlich, was translated in Japanese as bukimi prior to the publication of Mori’s paper. But MacDorman, who co-authored the definitive English translation of “The Uncanny Valley,” has his doubts: “There is nothing wrong with connecting Mori’s ideas to Freud,” he says. “But I don’t think Mori was inspired by him.”

In 2005 Mori began to get entangled with his study of the uncanny in much the same way that Freud had. In a somewhat puzzling note he sent to robotics conference, Mori wrote, “A dead person’s face may indeed be uncanny...[but] dead persons are free from the troubles of life, and I think this is the reason why their faces look so calm and peaceful.” These words came 35 years after the original publication of “The Uncanny Valley” and appear to suggest that what one finds uncanny evolves over time. MacDorman speculates that, in Mori’s case, this might be attributed to his age or development as a Buddhist. Here Weber’s point again rings true: Understanding the uncanny is neither an entirely subjective nor objective endeavor. Study it long enough, and eventually it makes a study out of you.

Evolving a Theory

But all along Mori hasn’t seen our avoidance of death as a consequence of repressed emotions the way Freud did. Instead he has understood it to be a mechanism we developed to keep ourselves safe. Nearly every hypothesis since has had this flavor. It has been suggested, for instance, that we avoid almost human figures because their peculiarities make them look sick, and we have developed an evolutionary mechanism for steering clear of pathogens. Another theory posits that we avoid figures with features slightly off from our own because they appear to be less-than-ideal mating material.

Ghazanfar rejects all of these hypotheses. “What is really going on is much simpler,” he says. He believes the uncanny valley response occurs because an animal—human or nonhuman—is evolutionarily inclined to develop an expectation of what members of its species should look like, a supremely important skill, as it lets the animal know with whom it can and cannot interact.

It’s easy to come up with new explanations, but hard to throw out the older ones.

In this sense, life-like robotic and computer-generated models occupy a weird middle ground in an animal’s mind: They are familiar enough for the animal to consider the possibility that they are of the same species, but strange enough that they don’t quite meet the expectation the animal has developed for members of its species. “Any face that violates that expectation is going to elicit the uncanny response,” Ghazanfar says.

There does appear to be some experimental evidence in support Ghazanfar’s theory. Studies with children have shown that at a very young age, babies do not react negatively to human-like robot as children grow older, such robots become more bothersome. This, Ghazanfar suggests, might be an indicator that infants have not yet developed a narrow expectation for what a human should look like. As of yet, however, he has not tested his theory explicitly. “It’s what I think, but the experiments with monkeys weren’t straightforward so I couldn’t address all those things,” he says, which puts him in much the same place as Freud, Mori, and others before him.

But even if Ghazanfar can prove that his theory is correct, it won’t necessarily disprove Freud or Mori. We just don’t know enough about the uncanny valley to be confident that it can be traced back to a single cause. And that’s always been one of the biggest difficulties studying the phenomenon: It’s easy to come up with new explanations, but hard to throw out the older ones. “Things can be uncanny because of perceptual mechanisms or more psychological mechanisms,” MacDorman says. “So I don’t think the uncanny valley is necessarily a kind of single phenomenon.”

The uncanny valley has shaped robotics design for the past 40 years in Japan. Computer generated characters in videogames and film are designed to avoid it. Yet a clear understanding of it—or even an agreed-upon definition—still escapes us. Ghazanfar hopes his research will help to address these questions someday soon, but for the time being we know little more for certain about its origins than we did when Ernst Jentsch first called our attention to it in 1906. Perhaps we should have heeded the German doctor’s cautionary clause as he began to broach the subject: “[If] one wants to come closer to the essence of the uncanny, it is better not to as what it is...”