

La valle del perturbante e la realtà virtuale. Posizioni e possibili linee di ricerca.

Elena Pasquinelli

Institut Jean Nicod – EHESS

Elena.Pasquinelli@ehess.fr

1. Introduzione

La “Valle del Perturbante” è una nozione all’interno di un semplice modello matematico introdotto dal robotico giapponese M. Mori [Mori, 1970] per descrivere l’evoluzione di una certa reazione, il “senso di familiarità”, di fronte ad artefatti robotici che presentano un livello crescente di realismo (nel senso di antropomorfismo o somiglianza con parti di esseri umani).

La prima parte del modello descrive la crescita del senso di familiarità in relazione all’incremento del realismo fino ad un punto di massima in seguito al quale si apre una valle di familiarità negativa; questa valle di non-familiarità è la Valle del Perturbante. Dall’altro lato della valle si alza un altro picco di familiarità positiva in corrispondenza con artefatti che riproducono perfettamente le fattezze umane e, al limite, con esseri umani veri e propri.

Il modello dà luogo ad una indicazione pratica indirizzata a creatori di robot e protesi, consistente nel consiglio di evitare di avvicinarsi troppo al primo picco della Valle: quando le condizioni proprie al secondo picco – perfetto realismo tanto per quel che riguarda l’aspetto fisico quanto per quel che riguarda il comportamento – non possano essere raggiunte, infatti, si rischia una caduta nella Valle del Perturbante. I robotici sono dunque invitati a volgere i loro sforzi verso artefatti meno realistici che non mirino a riprodurre le fattezze umane e verso i quali sia più facile sviluppare un senso di familiarità.

Al di là della sua applicazione al mondo della robotica, il modello contiene la seguente ipotesi, valida per le entità artificiali che simulano esseri viventi in generale: la ricerca del realismo si può rivelare controproducente, almeno nel caso in cui alcune precise condizioni non siano rispettate. Questa ipotesi generale è stata in effetti adottata da alcuni autori nell’ambito della critica di film d’animazione e della produzione di avatars e agenti in realtà virtuale, quindi dell’animazione in generale.

A. Jones, direttore dell’animazione in *Final Fantasy*, ad esempio, assicura che il tentativo di produrre delle repliche perfette di esseri umani può suscitare sensazioni inquietanti; i personaggi possono apparire grotteschi, facendoci l’effetto di un cadavere guidato come un burattino [Hiltzik, 2001]. Il gruppo di animatori Dream Works al lavoro sul film *Shrek* si è in effetti trovato nella condizione di dover limitare l’apparenza realistica della principessa Fiona per ovviare a simili spiacevoli effetti.

R. Ebert, noto critico cinematografico, adotta l’ipotesi alla base del modello della Valle del Perturbante per spiegare lo scarso successo di film d’animazione come “The Wild” e “Final Fantasy” [Ebert, 2006]. Nel primo dei due film, ad esempio, l’aspetto degli animali (pelliccia, peli e penne in particolare) e il loro comportamento (la sincronizzazione tra movimento delle labbra e articolazione dei suoni) sarebbero troppo ben fatti, e questo eccessivo realismo susciterebbe il sentimento che qualcosa non va e un senso di disagio. Nel secondo film l’animazione al computer è addirittura sfruttata per produrre personaggi che dovrebbero idealmente sostituire attori in carne ed ossa. Ebert attribuisce la caduta nella Valle del Perturbante al fatto che un realismo eccessivo distrugge la distorsione della realtà propria allo stile del disegno animato. La scelta di mantenere un aspetto fumettistico per personaggi dal movimento molto raffinato spiegherebbe dunque per contrasto il

successo dell'ultimo prodotto Pixar, "The Incredibles" [Canemaker, 2004].

Dal mondo del film d'animazione vengono dunque due indicazioni tese a evitare il rischio rappresentato dalla Valle del Perturbante: il rispetto dello stile fumettistico proprio alla natura dell'animazione, tanto per quel che riguarda la riproduzione di esseri umani che di altri organismi animali, e eventualmente la creazione di personaggi che non simulano entità esistenti nella realtà ma che possono comunque essere considerati realistici.

Un esempio di questo secondo genere di personaggi è Gollum in "The Lord of the Rings" [Ebert, 2004]: Gollum può essere considerato come uno dei personaggi più affascinanti e riusciti del film perché sembra una creatura reale ma non somiglia a nessuna creatura esistente. Gollum presenta infatti un livello di realismo assai elevato, un tipo di realismo che può essere associate con il secondo picco del modello di Mori, ma non simula esseri umani o animali.

Quel che conta, pare, non è dunque evitare il realismo in generale, ma il realismo nel senso della simulazione di entità umane o animali esistenti.

Alcune considerazioni preliminary possono essere estratte da questi esempi.

Primo, che l'animazione ha allargato i confini della Valle del Perturbante dal dominio della simulazione di esseri umani a quello di esseri viventi in genere (come nel caso degli animali di "The wild" o di Gollum di "the Lord of the Ring").

Secondo, che il termine "realismo" è usato nel contesto di queste discussioni almeno con due significati: realismo come simulazione di esseri umani o animali e una forma realismo senza simulazione (come nel caso di Gollum).

Terzo, il realismo sembra costituire un problema quando viola certe aspettative negli spettatori: l'aspettativa che lo stile di un cartone animato sia fumettistico e non realistico, le aspettative attivate dalla somiglianza con entità reali note allo spettatore, come esseri umani e animali.

Possiamo dunque considerare che in aggiunta al modello della Valle del Perturbante descritto da Mori (che costituisce una versione ristretta a robot antropomorfi, realistici nel senso di simulare le apparenze e il comportamento degli esseri umani), esiste un'ipotesi più ampia che deriva dal modello ristretto ma che si applica al problema più generale del realismo nella riproduzione di entità artificiali (entità materiali come i robot, fittizie come i personaggi di film o virtuali come gli agenti in realtà virtuale) che riproducono organismi esistenti o creature inedite. L'ipotesi ampia implica l'esistenza di una Valle del Perturbante in corrispondenza con certe condizioni che vanno da un elevato livello di realismo nella simulazione di entità esistenti alla violazione di aspettative.

Le considerazioni presentate in questa sezione verranno sviluppate in relazione alla chiarificazione di alcune nozioni introdotte da Mori, come la natura del sentimento che Mori descrive come "senso di familiarità", un termine del senso comune non ben specificato. Una migliore comprensione del fenomeno potrebbe essere acquisita specificando meglio le reazioni descritte da Mori.

Sembra infatti possibile discernere tre componenti nella sensazione di non-familiarità a partire dalle descrizioni offerte da Mori: emozioni legate a paura e disgusto (le più analizzate nella letteratura di robotica), un giudizio di tipo epistemico relativo alla credibilità dell'esperienza effettuata e una reazione di sorpresa negativa. Il problema della credibilità sembra essere generalmente connesso alla violazione di aspettative e alla violazione della coerenza. L'altro aspetto, connesso a emozioni di paura e disgusto, sembra essere più specificamente associato a esperienze che riguardano apparenze antropomorfe. Le emozioni di paura e disgusto potrebbero dunque costituire condizioni aggiuntive rispetto a quelle in causa nel problema della credibilità.

Possiamo dunque aggiungere una *quarta considerazione: che il modello della Valle del Perturbante presenta aspetti diversi, che questi aspetti sono tutti presenti nel modello ristretto ma possono anche essere analizzati separatamente nei termini di un problema generale di credibilità di*

entità artificiali e di un problema specifico di emozioni di perturbante legate a entità artificiali umanoidi.

2. Dimostrazioni della Valle del Perturbante

Visto l'interesse che potenzialmente può rivestire l'ipotesi generale sottesa al modello della Valle del Perturbante, sia dal punto di vista pragmatico (per la concezione di entità artificiali e robotiche) che teoretico (per il concetto di realismo e credibilità), è da auspicare che si individuino procedure di test adeguate a provare l'esistenza di una reazione del tipo Valle del Perturbante e che si possano individuare con precisione e misurare le condizioni atte a provocare tali reazioni nei diversi domini della robotica, dell'animazione, della realtà virtuale e eventualmente anche in altri tipi di rappresentazione.

Il modello della Valle del Perturbante è infatti ancora dibattuto. Alcuni autori, come [Hanson, 2003], considerano infatti il modello della Valle del Perturbante come non-scientifico. [Brenton, et al., 2005] sottolineano che le testimonianze aneddotiche quanto all'esistenza di reazioni non sono sufficienti a provare la correttezza del modello. Film e film d'animazione inoltre sono troppo complessi per poter attribuire accoglienza positiva o mancanza di successo al solo fatto di rispettare uno stile fumettistico piuttosto che alla bellezza estetica delle immagini o al valore della narrazione.

[MacDorman, et al., 2006b] ha recentemente fornito alcune prove quanto all'esistenza di reazioni di familiarità positiva e negativa analoghe al modello di Mori in soggetti esposti a fotografie di robot genericamente antropomorfi e veri e propri androidi (robot che riproducono il più possibile l'aspetto umano). Ai partecipanti all'esperimento è richiesto di valutare 31 immagini fotografiche lungo tre scale ciascuna da 9 punti: una scala di umanità (da molto meccanico a molto simile all'umano), una scala di familiarità (da molto strano a molto familiare), una scala di turbamento (da molto a poco strano e inquietante, con aggiunta della valutazione 0, per nulla strano e inquietante). [MacDorman, 2006a] fa infatti riferimento al sentimento descritto da Mori in giapponese col termine "bukimi" e legato alla Valle del Perturbante come "eeriness", ovvero un sentimento di turbamento provocato da qualcosa che appare inquietante, strano e eventualmente misterioso. Le fotografie presentano due tipi di contenuti: il primo gruppo presenta una transizione (attraverso un sistema di *morphing*) dalla foto di un robot vagamente antropomorfo (Qrio) a quella del robot super-realistico P. K. Dick (opera di Hanson), fino a una che ritrae l'autore di fantascienza in carne ed ossa; il secondo gruppo crea una transizione dall'immagine del robot antropomorfo Eveliee all'immagine dell'androide Repliee QI Expo alla foto del modello umano di Repliee. Per ambedue i gruppi, il risultato, quando la curva per la reazione di turbamento è tracciata in relazione alla curva di familiarità, lungo un asse che indica i diversi gradi di umanità, è un grafico che presenta lo stesso andamento del modello di Mori: una valle in corrispondenza con le immagini intermedie tra robot vagamente antropomorfo e robot androide, con un picco di turbamento in corrispondenza coi livelli più bassi di familiarità.

[Hanson, 2005b] considera però le reazioni di rigetto di fronte a robot umanoidi molto realistici come una pura illusione e sostiene che anche robot estremamente astratti o addirittura esseri umani dalle fattezze atipiche possono risultare estremamente perturbanti; d'altra parte, purché siano rispettate certe condizioni estetiche, qualunque livello di realismo può risultare attraente. Ci sono infatti vari modi di deviare dal realismo e non è scontato che le reazioni negative o positive siano in relazione col livello di realismo piuttosto che con la resa estetica. In un esperimento condotto con immagini analoghe a quelle impiegate da MacDorman, ma tali da risultare attraenti piuttosto che disturbanti, [Hanson, 2005b] mostra infatti che le reazioni del tipo Valle del Perturbante valgono solo per le immagini di controllo (non ritoccate), mentre le immagini ritoccate in modo da risultare attraenti non danno luogo a reazioni di turbamento e non sono valutate come inquietanti, e suggerisce il turbamento non dipende solo dal livello di realismo ma anche da considerazioni di tipo estetico. Per questo Hanson [Hanson, 2003, 2005a, 2005b; Hanson, et al., 2005] rivendica per la robotica l'obiettivo di creare

artefatti sempre più realistici nelle loro sembianze umane, purché siano rispettati convenienti criteri estetici.

Un altro esperimento condotto da MacDorman [MacDorman, 2006b] indica un'ulteriore fonte di variazione per le reazioni di turbamento: soggetti esposti a brevi filmati (dai 30 ai 60 secondi) di un'ampia gamma di robot umanoidi e androidi (tra cui un robot mobile, un braccio manipolatore, una testa androide, robot androidi completi, un essere umano) mostrano delle reazioni meno nette rispetto a quelle suscitate dalle immagini fotografiche, dato che la curva risultante dallo stesso tipo di valutazione adottata per le fotografie non presenta una sola Valle del Perturbante e che le curve estratte dalla valutazione della familiarità e dalla valutazione del turbamento sono pressoché speculari. [MacDorman, 2006b] suggerisce dunque che la percezione di artefatti realistici è influenzata da molteplici cause e che di conseguenza il perturbante occupa spazi diversi nel continuum che va dal realismo all'astrazione. Tra le cause del sentimento del perturbante sono inclusi la deviazione dalle norme di bellezza fisica, il senso di disgusto provocato dai segni indicatori di malattie e anomalie genetiche, rappresentazioni che evocano paura della morte o che ricordano l'inevitabilità della morte.

Tutte queste cause sono specialmente connesse al modello ristretto della Valle del Perturbante, poiché fanno riferimento a entità artificiali antropomorfe. MacDorman suggerisce infatti che i robot antropomorfi e gli androidi in particolare attivano negli esseri umani aspettative quanto al loro comportamento (incluse le interazioni sociali con esseri umani) che sono basate sull'esperienza che gli esseri umani hanno dei loro consimili. Per questa ragione i robot antropomorfi costituiscono uno strumento prezioso per lo studio dei comportamenti sociali umani e sono soggetti a reazioni negative da parte degli esseri umani quando le aspettative che attivano per certi dei loro aspetti non sono rispettate sotto altri punti di vista (come nel caso di un realismo imperfetto) [MacDorman, 2005a, 2005b, 2006a, 2006b; MacDorman, et al., 2006a, 2006b, 2006c]. La violazione delle aspettative in sé e per sé non è considerata come una condizione sufficiente per spiegare il senso di turbamento, inquietudine e stranezza provocato dagli artefatti umanoidi realistici: di fatto, non tutte le volte che vengono frustrate delle aspettative si fa esperienza di questo tipo particolare di sentimento [MacDorman, 2005b, 2006c]. Per questo motivo devono essere indicate cause aggiuntive che spieghino la specificità delle reazioni di fronte a entità artificiali umanoidi o simulazioni di esseri viventi.

Per poter sviluppare test atti a dimostrare l'esistenza di una Valle del Perturbante e individuare le cause di questo fenomeno o di fenomeni analoghi è comunque necessario compiere un passo preliminare, che consiste nel delucidare i diversi aspetti del modello e, preliminarmente, nell'analisi dello stato mentale associato alla Valle, che Mori descrive col termine di "non-familiarità" e che gli altri autori citati associano a un sentimento di turbamento, inquietudine, stranezza.

3. Senso di non-familiarità

La descrizione fornita da Mori del senso di familiarità può essere scomposta in diverse componenti, in parte connesse a reazioni emotive e in parte a giudizi epistemici.

Mori presenta infatti la Valle del Perturbante attraverso due esempi principali. Il primo caso esemplare è rappresentato da una mano protesica che riproduce tutti gli aspetti visivi di una mano reale; Mori sostiene che le persone che toccano questa mano rimangono scioccate dal fatto che è fredda e manca di tessuti molli: di colpo l'esperienza non è più familiare e suscita un senso di stranezza e turbamento o inquietudine. Il secondo esempio riguarda il movimento di arti protesici e in generale di robot umanoidi; Mori ipotizza che, nel caso più frequente in cui il movimento umano è reso solo imperfettamente, questi artefatti risulteranno scioccanti e tali da provocare un senso di orrore non solo al tatto ma anche alla vista.

Possiamo dunque estrarre le seguenti componenti del senso di non-familiarità:

- la componente di non-familiarità/stranezza indica la presenza di un giudizio epistemico relativo

alla *credibilità* dell'esperienza: in un certo senso, l'esperienza con l'entità perturbante non è credibile perché evoca un senso di falsità, di qualcosa che non può essere tale quale viene percepito. Le reazioni di shock possono essere assimilate a forti reazioni di *sorpresa*, e specificamente a reazioni di sorpresa negativa, ovvero di sorpresa associata con emozioni spiacevoli;

- la componente di non-familiarità/paura/repulsione: le emozioni negative che sia associano alla Valle del Perturbante includono un senso di inquietudine che sfocia nella *paura* e nella *repulsione*.

Ulteriore conferma della presenza di due componenti del senso di non-familiarità viene da esperimenti effettuati da MacDorman con immagini di teste di robot antropomorfi e androidi, nell'ambito della ricerca sulla possibilità che questo tipo di artefatti evocano negli esseri umani difese tipicamente dispiegate di fronte a condizioni che ricordano la morte e la sua inevitabilità [MacDorman, 2005b]. I commenti dei soggetti dell'esperimento fanno infatti riferimento a emozioni legate a paura e disgusto ("Immagine terrificante di donna", "Le prime due immagini erano inquietanti"; "Solo a guardarlo mi ha fatto venire la nausea"; "La donna faceva paura") e a un diverso tipo di impressioni connesse alla stranezza piuttosto che alla paura ("Strano"; "L'ho trovato bizzarro"; "Strana signora"). Strano, bizzarro sono aspetti del senso di non-familiarità differenti da paura e disgusto perché fanno riferimento a un giudizio che non è necessariamente connesso a emozioni negative ma piuttosto a reazioni di fronte a qualcosa di nuovo e inaspettato, a un'entità che appare finta o falsata e quindi non credibile.

È possibile avanzare un'ipotesi generale per spiegare la caduta di credibilità, generale nel senso che la diminuzione della credibilità può essere osservata in diversi tipi di esperienza e non solo nel caso in cui le entità percepite siano simili a esseri umani o viventi. Questa ipotesi si basa su osservazioni e risultati sperimentali ottenuti nell'ambito della realtà virtuale e non, e fa riferimento al ruolo giocato dalla coerenza e dalle aspettative nella percezione e agli effetti della violazione di queste condizioni.

Per quel che riguarda il secondo aspetto del senso di non-familiarità, un'ulteriore ipotesi suggerisce che le emozioni negative del tipo paura/disgusto possano essere associate alla caduta della credibilità in condizioni particolari, come nel caso in cui il contenuto dell'esperienza rappresenti volti umani o entità viventi in genere.

Queste considerazioni sono trasversali ai diversi tipi di esperienza, come esperienze mediate con entità fittizie e virtuali (cinema e realtà virtuale), esperienze non-mediate con entità artificiali e materiali (interazione con robot), esperienze non-mediate con entità reali e naturali.

4. Credibilità

L'opportunità di ricorrere al realismo per produrre personaggi e agenti credibili è stata valutata dai professionisti del settore tanto nell'ambito dei film d'animazione quanto in quello della realtà virtuale e di altri media più tradizionali, come i fumetti.

Un approccio classico al problema della credibilità è infatti rappresentato dai disegnatori Thomas e Johnston nel loro famoso testo "Disney: the illusion of life" [Thomas & Johnston, 1984]: la credibilità è concepita come un'illusione di realtà o vita, l'illusione che anche personaggi di film d'animazione caricaturali, non realistici e originali possono sembrare vivi. E questo proprio grazie a un uso adeguato di esagerazione e semplificazione nel disegno, nell'azione e nella rappresentazione di stati mentali cognitivi e emotivi. Lo stesso approccio è stato adottato per i fumetti e [McCloud, 1993] sottolinea il ruolo giocato dall'astrazione e dalla semplificazione nella resa di personaggi di fumetti: il lettore è in grado di identificarsi e può manifestare interesse grazie al loro livello di universalità e al fatto che sono presenti le loro caratteristiche essenziali. L'effetto di un requisito pratico dovuto alla necessità in animazione di produrre centinaia o migliaia di disegni, quello di usare immagini semplici e di cercare di astrarre il più possibile per estrarre le caratteristiche essenziali ad un personaggio, si è trasformato dunque in un requisito teorico per produrre personaggi più credibili.

A queste voci si aggiungono quelle dei ricercatori coinvolti nello “Oz project”, un progetto ospitato negli anni '90 dalla Carnegie Mellon University e teso a produrre agenti virtuali per opere teatrali interattive. [Bates, 1992, 1994] ad esempio accoglie la definizione disneyana di agenti credibili come di agenti che producono un'illusione di vita e permettono allo spettatore e utilizzatore di sospendere la sua incredulità e di avere una (falsa) impressione per cui il mondo e gli agenti con cui l'utilizzatore interagisce o che lo spettatore percepisce sarebbero reali anche se non somigliano a nulla che esista nella realtà.

I problemi suscitati dal realismo come simulazione di entità reali, e in particolare le aspettative che questo suscita negli spettatori e utilizzatori e che non necessariamente possono essere soddisfatte, danno quindi luogo all'indicazione generale di abbassare il livello di aspettative intrattenute, grazie a personaggi e agenti che non somigliano a nulla, come gli alieni [Scott-Reilly, 1992, 1995] o grazie al ricorso massiccio a tecniche di distorsione come la semplificazione e l'esagerazione. In ambedue i casi è accolto l'invito di Mori a mantenersi a distanza dal picco di realismo che precede la Valle del Perturbante. Uno spettatore o un utilizzatore con poche aspettative sul modo in cui personaggi e agenti dovrebbero essere fatti e si dovrebbero comportare sono più facili da soddisfare di spettatori con molte aspettative.

Ad ogni modo, sembra necessario aggiungere altre premesse che giustifichino l'invito ad abbassare il livello delle aspettative: che gli spettatori o utilizzatori hanno un certo numero di aspettative basate sull'esperienza e sulle loro conoscenze, che alcune di queste aspettative vengono attivate nel contesto dell'interazione con i mondi fittizi o virtuali e che il rispetto o la frustrazione di queste aspettative gioca un ruolo chiave nella credibilità di personaggi o agenti.

Queste premesse trovano in effetti un riscontro nell'osservazione e in alcuni risultati sperimentali e permettono di avanzare l'ipotesi che il problema della credibilità sia connesso al problema della coerenza e al rispetto delle aspettative attivate durante l'esperienza. La discrepanza tra il livello di realismo raggiunto per l'aspetto fisico e quello raggiunto per il comportamento, o tra livelli di realismo per aspetti diretti a differenti modalità sensoriali, o anche tra informazioni che non riguardano necessariamente lo stile più o meno realistico della rappresentazione, sembrano infatti essere responsabili di cadute di credibilità.

4.1 Coerenza tra comportamento e aspetto fisico

Mori suggerisce che il realismo, nel senso della simulazione di esseri umani, non dipende solo dall'aspetto fisico dell'artefatto, ma anche dal suo comportamento, in particolare il comportamento dinamico. Le marionette (Mori cita quelle del teatro giapponese Bunraku) sono indicate come esempi dell'importanza del movimento perché, benché non molto realistiche nelle loro apparenze fisiche, risultano simili a esseri umani grazie alle caratteristiche del loro movimento. Nel modello di Mori, la presenza di movimento ha l'effetto di cambiare la forma della curva della reazione di familiarità: picchi e valli ne risultano amplificati e anche lievi variazioni nella velocità o nel rapporto tra le caratteristiche di velocità e accelerazione sono in grado di provocare cadute nella Valle del Perturbante.

Esistono prove fin dagli anni '50 del fatto che la figura umana può venire identificata a partire dal suo movimento, semplicemente applicando dei punti luminosi alle articolazioni di un essere umano altrimenti invisibile allo spettatore e mostrando il comportamento dinamico dei punti luminosi contro uno sfondo scuro. Le considerazioni sul ruolo del movimento nel riconoscimento della forma umana possono essere estese al riconoscimento di emozioni, stile e genere [Pollik, 2004] e, al di là della percezione di altri esseri umani, a meccanismi generali di percezione di eventi dinamici [Viviani, 1990]. L'informazione sul comportamento dinamico influenza infatti anche la percezione della forma di oggetti, come dimostrato da alcuni esperimenti condotti da Viviani e colleghi: la variazione delle caratteristiche di velocità di punti luminosi che tracciano traiettorie circolari o ellittiche influenza la percezione della forma tracciata creando effetti illusori poiché il sistema percettivo tende a conformarsi

a una legge che connette le due caratteristiche e che è tipicamente rispettata dal movimento biologico [Viviani, 1989, 1997].

Anche nell'ambito della realtà virtuale, studi recenti attribuiscono un ruolo importante al comportamento dinamico nella valutazione di entità virtuali antropomorfe [Vynayagamoorly, 2005].

Nel caso di interazioni sociali con entità virtuali animate da umani (i cosiddetti "avatars") o animate dal computer (i cosiddetti "agenti") in ambienti virtuali la presenza di comportamento, e di comportamento dinamico in particolare, potrebbe rivestire un ruolo cruciale nel permettere agli utilizzatori quella che Dennett chiama "attribuzione di intenzionalità" [Dennett, 1996]. Dennett sostiene infatti che gli esseri umani tendono a proiettare la presenza di stati mentali come credenze e desideri in diversi tipi di entità, dagli animali alle creature artificiali anche semplici. Questa attribuzione di intenzionalità non può però essere attivata da qualunque entità naturale o artificiale, ma solo da entità che appaiono auto-guidate e quindi dotate di volizione. Ma di fatto solo entità che presentano una qualche forma di comportamento e in particolare di comportamento dinamico possono apparire auto o etero-dirette: finché si tratta di immagini statiche che non rappresentano alcuna reazione a eventi circostanti è impossibile pronunciarsi sulla loro natura volizionale.

[Nowak, et al., 2003] mostra che il fatto che un'entità virtuale sia animata dall'utilizzatore o dal computer non è rilevante nell'attribuzione di intenzionalità: in un esperimento condotto con avatars e agenti che presentano ciascuno tre livelli diversi di realismo (aspetto antropomorfo, aspetto astratto e un controllo senza rappresentazione visiva), le valutazioni del livello di presenza, co-presenza e presenza sociale da parte dei partecipanti non variano tra avatars e agenti; variano però coi tre livelli di realismo. È interessante notare che i due livelli estremi (realismo alto e controllo senza informazione visiva) producono gli stessi risultati, mentre un basso livello di realismo produce risultati peggiori. [Nowak, et al., 2003] suggeriscono che il tentativo di realismo attiva certe aspettative e che il basso livello di simulazione del livello intermedio frustra queste aspettative producendo un effetto negativo; al contrario l'alto livello di realismo rispetta queste aspettative, e l'assenza di tentativi di simulazione non produce aspettative e quindi non è suscettibile di produrre frustrazione.

Il ruolo delle aspettative in causa nella percezione di avatars virtuali è suggerito da un altro esperimento, descritto da [Garau, et al., 2003]. L'esperimento mostra infatti che quando aspetto e comportamento di figure virtuali antropomorfe sono in accordo quanto al livello rispettivo di realismo (anche nel caso di un basso livello di realismo), il risultato è migliore di quello ottenuto con un alto livello di realismo per una sola delle due condizioni. L'esperimento comprende quattro condizioni: due con livelli paragonabili di alto o basso realismo, tanto per l'aspetto fisico che per il comportamento dinamico, e due con livelli discrepanti di realismo, con l'aspetto fisico più realistico del comportamento dinamico e viceversa. Il comportamento in causa è rappresentato dall'espressività dello sguardo: nell'opzione realistica la fissazione dello sguardo dell'avatar è sincronizzata con i turni nella conversazione con l'utilizzatore, mentre nell'opzione a basso realismo la fissazione è casuale. Per ogni condizione vengono misurati attraverso un questionario diversi criteri di apprezzamento della relazione sociale, come il senso di co-presenza o il coinvolgimento dell'utilizzatore. I risultati indicano che un comportamento dinamico realistico produce effetti negativi quando associato a un aspetto fisico relativamente meno realistico, riducendo tutti i parametri tranne il coinvolgimento dell'utilizzatore, sul quale non pare avere effetto. Vale l'opposto per avatars dall'aspetto fisico più realistico.

L'importanza della coerenza tra livelli di realismo è confermata da [Tromp, et al., 1998] e da [Slater, et al., 2001].

Queste considerazioni tendono a confermare l'idea sottesa al modello della Valle del Perturbante, stando alla quale il realismo è un obiettivo rischioso, e l'idea supplementare stando alla quale l'implementazione di comportamenti dinamici in entità artificiali può giocare un effetto positivo quanto alla loro credibilità, ma anche incrementare i rischi di cadere nella Valle del Perturbante,

soprattutto nel caso in cui il livello di realismo impiegato per rappresentare i comportamenti dinamici non sia all'altezza di quello impiegato per l'aspetto fisico. All'altezza e non necessariamente superiore, perché quello che sembra contare di più è proprio la coerenza tra i diversi aspetti.

Sembra quindi plausibile ipotizzare che entità artificiali realistiche per l'apparenza fisica o per il comportamento attivino negli utilizzatori un certo numero di aspettative che investono ambedue gli aspetti dell'entità artificiale; l'effetto della violazione di queste aspettative sarebbe più potente dell'effetto di un elevato livello di realismo, anche quando si trattasse del realismo del comportamento dinamico, e consisterebbe in una rottura della relazione sociale tra agente o avatar e utilizzatore. L'utilizzatore non proverebbe necessariamente sensazioni di paura o disgusto, come descritto nel caso di robot antropomorfi e androidi, ma l'entità virtuale non attiverebbe nell'utilizzatore i comportamenti adeguati e desiderati per l'applicazione.

Se è vero che il sistema percettivo impiega l'informazione dinamica anche nel compito di riconoscimento della forma o aspetto fisico dell'entità percepita (come indicato dagli studi citati all'inizio del paragrafo sull'effetto del movimento nella percezione della forma) discrepanze tra informazione quanto al comportamento dinamico e informazione quanto all'aspetto fisico creerebbero un'ambiguità che renderebbe più difficile anche l'identificazione corretta dell'entità percepita.

Va comunque fatto notare che purtroppo non esiste uno standard comune quanto alla misurazione del livello di realismo in esperimenti condotti da diversi gruppi di studio. Così, quello che in un certo esperimento è indicato come "alto livello di realismo" può corrispondere ad un "basso livello di realismo" in esperimenti che si pongono obiettivi più ambiziosi. L'assenza di uno standard rende difficile comparare i risultati dei diversi esperimenti e impossibile considerare questi esperimenti come prove definitive dell'ipotesi proposta. La difficoltà è ulteriormente accresciuta dalla mancanza di standard anche per i criteri di valutazione: i testi di valutazione sono normalmente soggettivi e impiegano questionari diversi sviluppati localmente.

4.2 Coerenza tra aspetti diversi della psicologia dell'agente

Nell'ambito dell'animazione, il desiderio di maggiore realismo o illusione di realtà conduce i progettisti a cercare di produrre agenti "ricchi", ad esempio agenti dotati di emozioni e personalità. In aggiunta all'aspetto fisico e al comportamento motorio, questi agenti possono offrire prestazioni linguistiche, e le loro qualità fisiche, posture e azioni possono funzionare come segnali per stati emotivi, cognitivi e motivazionali e rappresentare una certa personalità. Si ritiene che questa "ricchezza" costituisca l'ingrediente fondamentale per creare entità virtuali o genericamente fittizie credibili [Thomas & Johnston, 1984]. I rappresentanti dello "Oz project", ad esempio, suggeriscono che dei personaggi credibili debbano mostrare, attraverso il loro aspetto e comportamento, di possedere una psicologia ricca [Bates, 1992, 1994; Mateas, 1999, 2002; Loyall, 1997], che includa gli stessi aspetti che ci aspettiamo di trovare in esseri umani naturali: personalità, emozioni, automotivazioni, cambiamento, relazioni sociali. La manifestazione di questi stati non deve necessariamente essere altrettanto sofisticata che negli esseri umani, ed anzi vale anche in questo caso l'invito a ricorrere a tecniche speciali di esagerazione e a tempistiche ben studiate per rendere le manifestazioni, ad esempio emotive, più chiare, isolate e quindi meglio riconoscibili.

Tuttavia, risulta difficile comprendere se i requisiti psicologici elencati siano quelli sufficienti e se sono tutti altrettanto necessari a produrre un effetto credibile.

Inoltre, l'aggiunta di psicologia realistica (almeno per quel che riguarda i suoi contenuti) può costituire una nuova fonte di rischio di caduta nella Valle del Perturbante perché crea nuove possibilità di conflitto. La possibilità di creare incoerenze è data dal fatto che gli esseri umani giudicano la personalità dei loro consimili sulla base di un certo numero di informazioni: il modo di parlare, di muoversi, eccetera. Uno studio condotto da [Nass, et al., 2000] suggerisce che la presenza di informazioni incoerenti che indicano la personalità ha conseguenze negative: agenti virtuali che

offrono informazioni contrastanti (contrasto tra comportamenti verbali e non-verbali) sono apprezzati di meno rispetto ad agenti che possono essere classificati in una certa categoria di personalità, gli utilizzatori tendono ad interagire di meno con essi e a ritenerli meno utili, meno divertenti e meno carismatici degli agenti coerenti. Non a caso, in effetti, nella realizzazione di personaggi di finzione il requisito di coerenza è tipicamente rispettato [Field 1994; Thomas & Johnston 1981].

L'apprezzamento negativo non coincide necessariamente con un giudizio di non-credibilità. Ciononostante, un elemento comune tra non-credibilità e incoerenza nella personalità è rappresentato dal fatto che i parlanti si volgono verso informazioni provenienti dal comportamento non verbale del proprio interlocutore per capire se l'interlocutore è sincero e può essere creduto o se le sue parole sono un inganno [Ekman & Friesen, 1974; Ekman, 1980]: l'incoerenza tra parole e comportamento non-verbale allerterebbe dunque rispetto alla presenza di inganno e falsità. Altre ragioni per preferire personaggi e agenti coerenti potrebbero essere costituite da alcuni vantaggi rappresentati dalla coerenza nel corso delle interazioni sociali, come la possibilità di predire correttamente quello che accadrà durante l'interazione [Fiske & Taylor, 1991], di ricordare meglio una persona [Cantor & Mischel, 1979], di alleggerire il carico cognitivo [Fiske & Taylor, 1991].

In ogni caso, sembra che, nel caso della credibilità, la coerenza e la sua violazione giocano un ruolo più generale, che non si limita all'accordo tra livelli di realismo e dunque allo stile con cui un agente o un personaggio sono rappresentati.

4.3 Coerenza tra modalità sensoriali

Come nel caso dei criteri per scoprire se il nostro interlocutore ci sta ingannando, altri tipi di violazioni della coerenza, come i conflitti intersensoriali, sembrano avere la funzione di allertare il soggetto dell'esperienza quanto alla presenza di un errore e di rendere possibile un giudizio epistemico di errore sulla sola base delle caratteristiche interne dell'esperienza.

Anche Mori fa riferimento ai conflitti intersensoriali quando illustra il caso esemplare della mano protesica che presenta un aspetto visivo realistico ma risulta fredda e rigida al tatto come suscettibile di provocare il sentimento di stranezza e inquietudine, con reazioni di sorpresa negativa. [MacDorman, 2005b] cita la possibilità che robot antropomorfi e androidi attivino e violino aspettative al livello intersensoriale, ma asserisce che la violazione della coerenza che si produce in questi casi non ha per effetto quel senso di inquietudine e turbamento che è tipico della Valle del Perturbante.

Esiste un certo numero di studi sulla percezione dedicati all'esperienza di conflitti intersensoriali e alla violazione della coerenza nella percezione che confermano gli effetti descritti da Mori e suggeriscono alcune ragioni per le reazioni negative di fronte ai conflitti percettivi.

I conflitti intersensoriali sono l'effetto dell'esposizione a stimoli multisensoriali e della combinazione di questi stimoli in un unico percolato finale. L'esperienza di conflitti intersensoriali è comunque piuttosto rara, anche in presenza di stimoli discordanti. Il sistema percettivo manifesta in effetti una tendenza a risolvere il conflitto in favore di soluzioni intermedie tra gli stimoli discordanti o in favore di dominanze unilaterali (che però non sono mai assolute) [Welch & Warren, 1981]. Questa tendenza suggerisce l'esistenza di una sorta di preferenza da parte del sistema percettivo per la coerenza. Nelle rare situazioni in cui il disaccordo non può essere risolto, il soggetto fa l'esperienza di un conflitto esplicito; l'esperienza del conflitto in questo caso si associa a reazioni di sorpresa e anche al giudizio che l'esperienza in causa appare impossibile, che quello che viene percepito non può essere vero, che deve essere sbagliato perché impossibile. Ad esempio, quando a certi muscoli degli arti viene impartita una vibrazione con strumenti adeguati, si ottiene una discrepanza tra l'informazione proveniente dai recettori presenti nel muscolo e nel tendine e i recettori presenti nell'articolazione [Goodwin, et al., 1972; Craske, 1977]. In questa occasione, il soggetto fa l'esperienza di un'illusione, come se l'arto che vien fatto vibrare si muovesse verso una certa posizione; nel caso in cui l'arto sia bloccato in una certa posizione dallo sperimentatore, il movimento esperito è tale da sorpassare i limiti

anatomici delle articolazioni. In tal caso il movimento è immediatamente identificato come impossibile dai soggetti dell'esperimento, che infatti commentano che quello che percepiscono non può essere vero: *"il braccio si sta rompendo"*, *"è piegato all'indietro, non può essere dove lo sento"* e che il loro braccio sembra trovarsi *"in due posti allo stesso tempo"*. I soggetti sono quindi messi in condizione di esprimere un giudizio epistemico a proposito della loro esperienza (secondo il quale c'è per forza un errore) sulla base delle caratteristiche dell'esperienza stessa.

In altre parole, l'esperienza di un conflitto ha l'effetto di allertare il soggetto quanto alla presenza di un errore e di metterlo in guardia quanto alla credibilità delle componenti dell'esperienza stessa. Per questo motivo le violazioni della coerenza tra modalità sensoriali potrebbero giocare un ruolo così negativo sulla credibilità di un'esperienza.

Le violazioni della coerenza percettiva non si limitano però ai casi di conflitto intersensoriale; ne sono altri esempi i paradossi percettivi e in particolare l'esperienza con figure impossibili e ambigue (come il triangolo di Penrose, una rappresentazione possibile quando vengano seguite le regole per il disegno bidimensionale, ma impossibile e si seguono le regole del disegno tridimensionale; o come la figura del coniglio-anatra che appare ora come un animale ora come un altro). Come nel caso dei conflitti intersensoriali, la violazione della coerenza provoca una reazione di sorpresa accompagnata dalla sensazione che qualcosa non va. E come nel caso dei conflitti, l'esperienza percettiva risulta ambigua e può essere interpretata in due modi diversi. È da notare che le due interpretazioni sono incompatibili: il soggetto della percezione sembra non avere altra scelta che accedere ad un aspetto alla volta, anche quando è a conoscenza della particolarità della figura esperita.

Il caso dei paradossi percettivi permette dunque di generalizzare nel modo seguente: in tutti i casi in cui il soggetto si trova ad essere indeciso tra due interpretazioni di dati percettivi che non sono mutuamente compatibili, l'esperienza risulta bizzarra, strana, suscita sorpresa e eventualmente la sensazione di qualcosa di sbagliato (e quindi non credibile), quando i dati percettivi possano essere ricondotti a qualche forma di chiara violazione della coerenza.

La presenza di reazioni di sorpresa nei casi descritti permette di costruire un legame tra le violazioni della coerenza e le violazioni di aspettative. Le reazioni di sorpresa sono infatti tipicamente associate alla presenza di aspettative. Nel caso di aspettative frustrate è violata la coerenza tra due gruppi di aspettative; queste aspettative possono derivare da credenze o conoscenze acquisite in diverso modo e vengono riattivate dal contesto percettivo. Nel conflitto intersensoriale della mano protesica descritto da Mori, ad esempio, i dati visivi evocano certe aspettative per la modalità tattile che non sono però rispettate dall'effettiva esperienza tattile.

Possiamo quindi avanzare l'ipotesi che una delle ragioni per cui esperienze non coerenti - in cui aspettative di vario genere vengono violate (aspettative quanto al comportamento di altre modalità sensoriali, aspettative quanto al livello di realismo degli enti percepiti) - risultano non credibili risiede nel fatto che le violazioni della coerenza sono un campanello di allarme epistemico che segnala al soggetto la presenza di errori o falsità.

Un'altra ragione possibile per spiegare l'effetto negativo delle violazioni della coerenza è suggerita da alcuni studi dedicati al valore della coerenza a livello dei comportamenti adattivi. Di fatto, quando un soggetto si trova di fronte a una ambiguità o a un conflitto, gli risulta difficile agire adeguatamente, poiché la percezione non guida le sue azioni verso un obiettivo chiaramente identificato.

[Stein & Meredith, 1993] ad esempio sottolineano il valore positivo dell'integrazione a livello centrale di informazioni multisensoriali provenienti dal tatto, dalla visione e dall'udito con i centri dedicati alla pre-programmazione del movimento (i loro studi si concentrano sul funzionamento del colliculus superiore nel gatto). L'integrazione multisensoriale abbassa infatti la soglia di attivazione dei neuroni di ogni modalità sensoriale e permette di dare luogo a risposte motorie più rapide e quindi più

valide da un punto di vista adattativo. Quando però l'informazione è incoerente, l'effetto sull'azione diventa distruttivo: l'azione risulta infatti inibita o diretta sull'obiettivo sbagliato.

E' plausibile ipotizzare che l'inibizione dell'azione costituisca un effetto di violazioni di coerenza in generale a causa dell'ambiguità delle informazioni disponibili e della difficoltà a programmare un piano d'azione conseguente, e che questa inibizione sia percepita come negativa dal soggetto e quindi associata a emozioni spiacevoli in virtù del suo valore adattivo negativo. Le emozioni di disgusto e paura non sono comunque conseguenza diretta della violazione di coerenza e aspettative, e sembrano richiedere condizioni aggiuntive più specifiche, legate alla percezione di entità antropomorfe o simili a esseri viventi. Il senso di non-familiarità descritto da Mori sarebbe dunque presente nella sua forma completa solo in queste situazioni speciali, ma l'idea di una Valle negativa di non-credibilità, stranezza, sorpresa spiacevole sarebbe estesa a condizioni molto più generali il cui contenuto non fa necessariamente riferimento ad agenti o avatar che imitano esseri viventi.

Il ruolo giocato dalla coerenza e dal rispetto delle aspettative illustra comunque la possibilità che altre caratteristiche dell'esperienza, al di qua del realismo, siano da prendere in considerazione per rispondere all'obiettivo della credibilità. Le violazioni di coerenza e delle aspettative, ad esempio, sono fattori a cui fare molta attenzione dal momento che si sono rivelate fatali per la credibilità. Le indicazioni tese ad abbassare il livello delle aspettative attraverso tecniche e stili speciali nella rappresentazione o attraverso la creazione di entità originali vanno in questa direzione.

Non è insomma detto che solo agenti e personaggi realistici siano suscettibili di essere credibili, e anzi, come abbiamo visto nell'introduzione e come testimoniato dai professionisti di diversi settori, personaggi come il Gollum di "The Lord of the Rings" o alieni di varia natura risultano credibili senza simulare nessun essere vivente esistente. Se riprendiamo le considerazioni sviluppate nei paragrafi precedenti, il realismo sembra dunque non solo costituire una notevole fonte di problemi per la creazione di personaggi e agenti credibili, ma anche un requisito non necessario per produrre credibilità.

5. Il perturbante come repulsione e paura

La nozione di non-familiarità e perturbante intesa come un senso di inquietudine connesso con esperienze narrative o con certe situazioni della vita quotidiana è al centro di un saggio classico di Freud [Freud, 1919]. La parola "perturbante" è in questo caso la traduzione del tedesco *Unheimlich*. Il caso esemplare analizzato da Freud è quello della novella di Hoffmann, Il Mago Sabbiolino, in cui un automa di nome Olimpia viene scambiato dal protagonista del racconto per un vero essere umano. Il senso di perturbante è attribuito al fatto che qualcosa che è familiare al soggetto (*Heimlich*) ma che avrebbe dovuto rimanere nascosto (ancora un significato della parola *Heimlich*) è venuto allo scoperto; si tratterebbe di un'istanza del meccanismo noto come "ritorno del rimosso". Nel caso del Mago Sabbiolino in particolare circola secondo Freud con insistenza un complesso di castrazione rappresentato dalla paura di perdere gli occhi e dagli occhi naturali di Olimpia, sottratti dal Mago Sabbiolino ai bambini; il complesso rimosso del lettore verrebbe rievocato dal racconto e dà luogo al sentimento del perturbante. Bambole, Doppelgänger, automi sarebbero particolarmente efficaci nel produrre il senso del perturbante perché evocano epoche remote dell'infanzia di ognuno di noi, in cui dominavano pensieri di tipo animistico secondo i quali la vita permea qualunque entità e ognuno di noi ha una sorta di Doppelgänger sotto forma di anima. Freud nega però che a produrre il sentimento di perturbante sia l'incertezza cognitiva, ad esempio quella concernente il fatto che una certa creatura sia realmente animata o sia solo un automa o anche l'incertezza tra vita e imitazione della vita o rappresentazione.

L'idea dell'incertezza cognitiva è stata recentemente riproposta da [Ramey, 2005, 2006]. Ramey avanza l'ipotesi che l'effetto di perturbante non sia specifico ai robot antropomorfi ma sia parte di una classe più ampia di stati cognitivi e percettivi di incertezza rispetto ai confini tra diverse

categorie, come per i confini tra uomo e robot.

L'ipotesi dell'incertezza cognitiva in generale in effetti non sembra essere sufficiente a spiegare la specificità dei sentimenti di paura e disgusto, perché, come abbiamo visto, molte forme di deviazione dalle aspettative e violazioni di coerenza legate a situazioni ambivalenti possono essere prodotte senza dare luogo necessariamente alla forma completa di senso di non-familiarità descritta da Mori, ma solo a cadute della credibilità e sorpresa.

[MacDorman, 2006a] avanza l'ipotesi che i robot antropomorfi, in particolare quando il loro realismo non è perfetto (ad esempio al livello del movimento) evocano pensieri di morte, come fanno gli zombi. In generale, viene infatti ipotizzato che i robot antropomorfi evocano lo stesso tipo di aspettative che sono attivate dall'interazione con esseri umani, ma che robot che non siano perfetti nella loro somiglianza con esseri umani hanno l'effetto di ricordare la morte. I pensieri di morte a loro volta evocano difese, conscie o inconscie, come quelle descritte e misurate dalla teoria del Controllo della Paura. I risultati di un test comportamentale basato su questa teoria e descritti da [MacDorman, 2005b] sembrano confermare la presenza di questo tipo di difese in soggetti esposti a robot antropomorfi. Visto che più cause possono influire sul medesimo fenomeno, altre ipotesi avanzate da [MacDorman, 2006a] riguardano la presenza di indicatori di malattia come patologia o malattia genetica o anche la presenza di caratteristiche che evocano altri tipi di paure più o meno legate alla morte, come la paura di essere sostituiti da una sorta di Doppelgänger, la paura di entità viventi senza anima (come l'automa Olimpia), o evocare immagini come quelle di un corpo smontato in pezzi (come nella favola di Giovannin Senza Paura). In tutti questi casi l'evocazione della morte può dare luogo a segnali di pericolo.

[Hanson, 2003] presenta una serie di studi neurofisiologici che dovrebbero confermare l'idea per cui l'effetto di turbamento non è legato al realismo in sé ma è piuttosto il risultato del funzionamento di un sistema di allarme che riconosce entità antropomorfe e si attiva in corrispondenza con condizioni di pericolo potenziale, come nel caso in cui non siano presenti i segni considerati tipici della buona salute. Viene proposta anche una seconda ipotesi, secondo la quale in presenza di entità antropomorfe lo stesso sistema produce aspettative di verosimiglianza sempre più avanzata; se le aspettative sono confermate può iniziare una relazione sociale, ma quando queste aspettative non vengono rispettate viene attivata una risposta di paura. Inoltre, come abbiamo visto precedentemente, la manipolazione delle caratteristiche estetiche dei robot al fine di produrre entità più attraenti pare avere un effetto positivo nella riduzione della Valle del Perturbante, e suggerisce che anche la presenza di caratteristiche che sono valutate negativamente da un punto di vista estetico possa giocare un ruolo nell'attivare segnali di pericolo legati alla malattia come deviazione dalla norma [Hanson, 2005b].

Le ipotesi avanzate da Hanson e MacDorman si basano sulla specificità della figura umana e delle incertezze e paure legate a questa immagine e non ad altre. Addirittura il volto potrebbe costituire l'elemento decisivo. L'idea stessa della somiglianza tra umani e robot non è scontata, e le similitudini rilevate da soggetti di un esperimento condotto da Ramey appartengono piuttosto a componenti del corpo che al volto [Ramey, 2005, 2006]. Potrebbero dunque essere le dissimilitudini rilevate nel volto dei robot antropomorfi rispetto al loro modello umano a scatenare reazioni di turbamento.

5.1 Fisiologia e patologia del riconoscimento di volti

I volti umani e le loro rappresentazioni costituiscono in effetti oggetti speciali per la percezione: non solo aree specifiche del cervello sono dedicate al loro riconoscimento e alla loro identificazione, ma queste aree sono attivate anche da rappresentazioni astratte che presentano un numero minimo di dettagli. L'artista di fumetti McCloud offre un'illustrazione molto evocativa dell'esistenza di una predisposizione forte verso l'identificazione di facce, anche a partire da figure molto semplici: se disegniamo una serie di figure chiuse dai confini anche molto irregolari e poi piazziamo una piccola forma circolare con un pallino al centro (tale da ricordare un occhio stilizzato) in un punto qualunque

all'interno della traccia, questa apparirà immediatamente come un volto, bizzarro, ma comunque un volto.

La neurofisiologia e patologia della percezione di volti possono dunque costituire un valido aiuto nella ricerca di spiegazioni scientifiche della reazione di paura e disgusto associata a robot antropomorfi imperfettamente realistici. [Hanson, 2003] propone in effetti una rassegna di studi sulla percezione di facce e indica in particolare negli studi di neuro-immagini la fonte privilegiata di informazione per la comprensione del fenomeno della Valle del Perturbante e l'individuazione delle cause che provocano reazioni di perturbante.

Anche la patologia può però provvedere prove quanto alla specificità della percezione di volti umani e può di conseguenza aiutare a individuare alcune delle condizioni che fanno sì che la percezione di un volto diventi inquietante, disturbante e addirittura evochi paura e disgusto. I fenomeni classificati come sindromi di misidentificazione in particolare presentano alcune affinità con le sensazioni della Valle del Perturbante.

La *Sindrome di Capgras* [Ellis & Lewis, 2001] consiste in un disturbo caratterizzato dalla credenza per cui un parente stretto sarebbe stato rimpiazzato da un impostore, robot o alieno, come avviene in diversi film e romanzi aventi per tema l'invasione di alieni che si sostituiscono a umani e ne prendono il posto, ingannando anche i famigliari (vedi "L'invasione degli ultracorpi" [Finney, 1955]). L'errore di identificazione sembra essere legato a una perdita della risposta emotiva di fronte a volti familiari visto che sono assenti le risposte autonome differenziali alla percezione di volti familiari e non familiari; allo stesso modo, gli alieni sostituiti ai veri esseri umani vengono riconosciuti grazie alla loro mancanza di emozioni. E' proprio la mancanza di emozioni a creare un sentimento di paura e impostura.

Un altro fenomeno legato alla Sindrome di Capgras è la cosiddetta *Sindrome dei Doppi* [Christodouloun, 1978], in cui il soggetto crede nell'esistenza di un doppio di sé stesso o Doppelgänger, avente lo stesso aspetto fisico ma una diversa personalità. Anche questo è un tema ricorrente di film e racconti ("Doppelgänger" [Anderson & Anderson, 1969], "Mirror image" [Sterling, 1959]).

Infine, nella *Sindrome di Fregoli* [Ellis, et al., 1994] il soggetto ha il convincimento di essere perseguitato da una persona in grado di prendere diversi aspetti; ritiene quindi che molte delle persone che incontra non sono che degli impostori.

L'interesse rappresentato da queste sindromi consiste nel mostrare casi di percezione in cui l'apparenza non è più legata all'identità, casi di percezione di impostori che creano un sentimento di paura, rigetto e disagio nel soggetto della percezione. Queste sindromi sono attribuite a disfunzionamenti nei meccanismi di percezione delle facce, dovuti a danni cerebrali o altri disordini neurologici. Un'ipotesi per spiegare la paura e il disgusto o disagio provocati dalla percezione di figure umane imperfette potrebbe dunque essere la seguente: la visione di un volto umano o una sua rappresentazione attiva specifiche aree cerebrali, anche se la rappresentazione è astratta o incompleta; tuttavia, quando la figura presenta caratteristiche particolari, ad esempio tali da contraddire i dati che segnalano la presenza di un volto umano, o di un volto umano familiare, o di un volto umano "normale", il meccanismo di riconoscimento delle facce comincia a malfunzionare, in un modo che è analogo al malfunzionamento che produce le sindromi di misidentificazione. Il volto è dunque percepito in un certo senso come appartenente ad un impostore e l'impostore associato a paura e rigetto.

6. Conclusioni

Il modello della Valle del Perturbante, nelle sue versioni ristretta e allargata, apre diverse linee di ricerca possibili.

Nella sua versione ristretta il modello consiste in una curva che descrive le reazioni degli esseri

umani di fronte a robot antropomorfi; questa curva richiede ulteriori ricerche sperimentali perché la sua esistenza e forma vengano confermate.

Nella sua versione allargata, il modello della Valle del Perturbante presenta un certo numero di testimonianze aneddotiche in diverse applicazioni di interazioni mediate e alcune conferme sperimentali, difficili però da confrontare tra di loro per la mancanza di standard di misura. Gli elementi che compongono il modello allargato includono agenti e personaggi reali, fittizi o virtuali che presentano una qualche forma di realismo. È plausibile pensare che anche la rappresentazione di oggetti e non solo di entità viventi produca una parte degli effetti descritti come Valle del Perturbante, specificamente la caduta della componente di credibilità. Questa ipotesi non è ancora stata sottoposta a ricerca sperimentale.

Qualunque attività sperimentale concernente la Valle del Perturbante richiede in ogni modo una chiarificazione preliminare quanto dei concetti coinvolti, in particolare per quel che riguarda gli stati psicologici e le reazioni effettivamente attivate negli esseri umani, il significato e la quantificazione del realismo, le entità significative per scatenare le reazioni proprie al modello.

Anche il termine “realismo” del resto si è rivelato ambiguo. Da una parte il realismo è l'effetto della simulazione del mondo reale, la riproduzione che si vuole il più possibile fedele alle caratteristiche fisiche o percettive, ed eventualmente anche comportamentali e psicologiche di entità che esistono nel mondo reale; dall'altra parte il realismo viene inteso come uno stato mentale, un sentimento basato su una forma di inganno che fa sì che mondi e entità virtuali o fittizi appaiano reali quanto il nostro mondo, pur senza presentare somiglianze con questo.

Nel primo significato, una rappresentazione viene descritta come più o meno realistica nel senso di presentare più o meno somiglianze con le proprietà di entità esistenti. Non è facile quantificare questa forma di realismo, e lo prova il fatto che si parli di realismo tanto per immagini fotografiche, che per disegni in bianco e nero più o meno stilizzati (come nel caso dei fumetti), per opere pittoriche cosiddette “fotorealistiche”, per modelli virtuali che presentano livelli di dettaglio più o meno elevati. Questi esempi riguardano solo l'aspetto percettivo visivo delle rappresentazioni; ma si possono avere rappresentazioni uditive, come nel caso di radio o film, e anche tattili, come nel caso di ambienti virtuali multisensoriali. Non necessariamente le condizioni di realismo visivo valgono per le altre modalità sensoriali. Inoltre il realismo può riguardare la rappresentazione di stati mentali emotivi o di altro genere; in questo caso agli aspetti percettivi della manifestazione delle emozioni si aggiungono aspetti puramente narrativi concernenti il tipo di emozioni o altri stati mentali riprodotti e il modo di interagire tra di loro. Questo tipo di realismo è richiamato anche nel caso di media che non si basano sulla percezione e sull'interazione, ma solo sulla narrazione, come i libri. Infine: in realtà virtuale il problema della simulazione si pone in maniera duplice: sono le caratteristiche fisiche o quelle percettive che devono essere simulate per ottenere il miglior risultato?

In secondo luogo il realismo è stato presentato come uno stato mentale e non come una caratteristica intrinseca degli enti simulati: una rappresentazione è realistica nel senso che evoca un senso di realtà e, per entità particolari, di vita. Siccome l'entità in questione non esiste nel mondo reale, questo stato mentale è considerato un'illusione. L'“illusione di realtà” è un termine tutt'altro che chiaro, sia da un punto di vista concettuale che psicologico. In che cosa consista questa sensazione non è facile dirlo. Potrebbe essere intesa come una forma di illusione di non-mediazione, come l'illusione dalla parte dell'utilizzatore o dello spettatore di essere “là”, nel mondo del film o nel mondo virtuale. Ma non è evidente che utilizzatori e spettatori possano compiere questo genere di errore. Non è scontato infatti presumere che gli esseri umani possano dimenticare la loro posizione di spettatori e utilizzatori, e anzi, il tipo di emozioni e reazioni manifestate anche di fronte a spettacoli particolarmente convincenti non sembra tale da confermare l'esistenza di una vera e propria obliterazione della consapevolezza della presenza di un medium. Di fatto, da molto tempo nessuno corre più via dalle sale cinematografiche in preda al terrore per la vista di un treno che avanza e nessun aviatore è mai stato

ingannato da un simulatore di volo al punto da cercare di buttarsi col paracadute dopo qualche errore fatale [Stoffregen, et al., 2003]. Sembra piuttosto che, nel migliore dei casi, spettatori e utilizzatori si prestino a fare come se quello che esperiscono fosse vero, pur sapendo che non lo è, e a mettere in campo una serie di azioni adeguate alla situazione che sembra vera e al medium in causa (così, nel caso di un medium interattivo come la realtà virtuale, gli utilizzatori potranno effettivamente cercare di esplorare e interagire, mentre nel caso di media passivi come il cinema si limiteranno a mettere in gioco le loro reazioni emotive e risorse attenzionali). Per questo motivo abbiamo introdotto il concetto di credibilità, uno stato mentale anche questo, ovvero un giudizio dalla parte dello spettatore o utilizzatore accompagnato da reazioni emotive e comportamentali, che però non richiede necessariamente l'obliterazione della propria condizione di spettatore e utilizzatore. La credibilità potrebbe semplicemente richiedere allo spettatore di accettare di giocare un gioco di far finta, come quello dei bambini che fingono di tenere in mano una spada quando invece maneggiano un bastone.

Il rispetto delle aspettative attivate permetterebbe allora agli adulti spettatori e utilizzatori di entrare in un gioco in cui fare come se le entità e i mondi esperiti fossero veri. Abbiamo visto infatti che il rispetto delle aspettative attivate e della coerenza costituisce un elemento fondamentale per mantenere la credibilità e non cadere in una Valle di esperienze che appaiono bizzarre al punto da allertare spettatori e utilizzatori alla presenza di un qualche errore o di esperienze che presentano elementi troppo discordanti per non suscitare un senso di disagio, di impossibilità e di sorpresa negativa. La letteratura direttamente dedicata al problema della credibilità non è vastissima, ma abbiamo visto che l'idea generale di una Valle negativa si ripresenta sia che si parli di presenza, co-presenza, divertimento o interesse. A produrre la caduta non è necessariamente un livello di realismo elevato o insufficiente, ma spesso la presenza di discrepanze e violazioni delle aspettative che possono essere quindi intese giocare un ruolo più generale.

La Valle negativa presenta componenti diverse, identificate da Mori sotto il termine generico di "senso di familiarità". Alle componenti più epistemiche della credibilità si possono associare componenti emotive caratterizzate da paura e repulsione, almeno quando certe condizioni specifiche siano soddisfatte, come nel caso di entità robotiche, fittizie o virtuali che simulano esseri umani in modo imperfetto. Tra le ragioni che rendono speciali le simulazioni antropomorfe abbiamo indicato la natura speciale delle interazioni sociali e della percezione di volti.

La Valle del Perturbante appare dunque come un fenomeno molto ricco, tale da interessare non solo i domini applicativi della robotica e della realtà virtuale o della finzione, ma capace di suscitare quesiti profondi e indicare direzioni di ricerca anche all'investigazione filosofica e psicologica.

Bibliografia delle opere citate

- Anderson, S., & Anderson, G. (1969). *Doppelgänger* (R. Parrish, Director).
- Bates, J. (1994). *The Role of Emotion in Believable Agents* (CMU-CS-94-136). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Bates, J., Loyall, A. B., & Reilly, W. S. N. (1992). *An Architecture for Action, Emotion, and Social Behavior*. Paper presented at the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World, S.Martino al Camino, Italy.
- Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). *The uncanny valley: does it exist?* Paper presented at the 19th British HCI Group Annual Conference: workshop on human-animated character Interaction, Edinburgh.
- Canemaker, J. (2004, October 3, Sunday). Part Human, Part Cartoon: A New Species. *New York Times*, pp. 13.
- Cantor, N., & Mischel, W. (1979). Prototypes in person perception. *Advances in Experimental Social Psychology*, 12, 3-52.
- Cavell, S. (1986). *The uncanniness of the ordinary*. Unpublished manuscript, Stanford University.
- Christodouloun, G. N. (1978). Syndrome of subjective doubles. *Am J Psychiatry*, 135(2), 249-251.
- Craske, B. (1977). Perception of impossible limb position induced by tendon vibration. *Science*, 196, 71-73.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, Mass.: MIT Press.
- Dennett, D. C. (1996). *Kinds of minds : toward an understanding of consciousness*. New York, N.Y.: Basic Books.
- Ebert, R. (2004, January 11, 2004). Movie Answer Man. *Chicago Sun Times*.
- Ebert, R. (2006, April 14, 2006). The wild. *Chicago Sun Times*.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social*

Psychology, 29, 288–298.

- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38(2), 270–277.
- Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends in cognitive sciences*, 5(4), 149-156.
- Ellis, H. D., Whitley, J., & Luaute, J. P. (1994). Delusional misidentification. The three original papers on the Capgras, Fregoli and intermetamorphosis delusions. (Classic Text No. 17). *Hist Psychiatry*, 5(17), 117-146.
- Field, S. (1994). *Screenplay: The foundations of screenwriting*. New York: Bantam Doubleday Dell.
- Finney, J. (1955). *The Body Snatchers*. New York: Dell.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.
- Freud, S. (1955). *The Uncanny [Das Unheimliche trans. James Strachey 1919]*. (Vol. 17). London.
- Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). *The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment*. Paper presented at the CHI 2003: Proceedings of the SIGCHI conference on Human factors in computing systems.
- Goodwin, G. M., McCloskey, D. I., Matthews, P. B. C. (1972). Proprioceptive illusions induced by muscle vibration: contributions by muscle spindles to perception? *Science*, 175, 1382-1384.
- Gregory, R. L. (1973). The confounded eye. In R. L. Gregory & E. H. Gombrich (Eds.), *Illusion in Nature and Art*. London: Duckworth.
- Gregory, R. L. (1997). Visual Illusions Classified. *Trends in Cognitive Sciences*, 1(5), 190 -194.
- Hanson, D. (2003). Investigating the neural basis of the uncanny valley.
- Hanson, D. (2005). *Expanding the aesthetic possibilities for humanoid robot*. Paper presented at the IEEE-RAS international conference on humanoid robot, Tsukuba, Japan.
- Hanson, D. (2005b). *Exploring the Aesthetic Range for Humanoid Robot*. Paper presented at the Toward Social Mechanisms of Android Science
- Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., Fernandez, R., & Stephanou, H. (2005). *Upending the uncanny valley*. Paper presented at the Proceedings of the Twentieth
- Hiltzik, M. A. (2001, May 8, 2001). Synthetic Actors Guild. *Los Angeles Times*.
- Isbister, K., & C., N. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53(2), 251 - 267.
- Loyall, A. B. (1997). *Believable Agents: Building Interactive Personalities* (CMU-CS-97-123). Pittsburgh, PA: School of Computer Science Computer Science Department Carnegie Mellon University.
- MacDorman, K. (1999). Grounding symbols through sensorimotor integration. *JourNal of the Robotics Society of Japan*, 17, 20-24.
- MacDorman, K. F. (2005a). *Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it?*
- MacDorman, K. F. (2005b). *Mortality salience and the uncanny valley*. Paper presented at the IEEE-RAS International Conference on Humanoid Robot, Tsukuba, Japan.
- MacDorman, K. F. (2006a). Introduction to the special issue on android science. *Connection Science*, 18.
- MacDorman, K. F. (2006b). *Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley*. Paper presented at the CCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, Vancouver, Canada.
- MacDorman, K. F., & Ishiguro, H. (2006a). Opening Pandora's uncanny box: Reply to commentaries on "The uncanny advantage of using androids in social and cognitive science research.". *Interaction Studies*, 7(3), 361-368.
- MacDorman, K. F., & Ishiguro, H. (2006b). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3).
- MacDorman, K. F., & Ishiguro, H. (2006c). *Toward social mechanisms of android science*. Paper presented at A CogSci 2005 workshop. *Interaction Studies*, 7(3).
- Mateas, M. (1999). An Oz-centric review of interactive drama and believable characters. In M. J. Wooldridge & M. Veloso (Eds.), *Artificial intelligence today: Recent trends and developments* (pp. 297-398). NY: Springer.
- Mateas, M. (2002). *Interactive Drama, Art, and Artificial Intelligence* (CMU-CS-02-206). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- MCCloud, S. (1993). *Understanding comics*.: Kitchen Sink press.
- Mori, M. (1970). The Uncanny Valley. *Energy*, 7, 33-35.
- Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: Researching conversational agents., *Embodied conversational agents* (pp. 374-402). Cambridge, Mass.: MIT Press.
- Nowak, K., & Biocca, F. (2003). Understanding the influence of agency and anthropomorphism on copresence, social presence and physical presence with virtual humans. *Presence: Teleoperators and Virtual Environments*, 12(5), 481-494.
- Pollik, F. (2004). *The search for the uncanny valley*. Paper presented at the Workshop at the Konrad-Lorenz-Institute for Evolution and Cognition Research, Altenberg Austria.
- Ramey, C. H. (2005). *The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robot*. Paper presented at the IEEE-RAS international conference on humanoid robot, Tsukuba, Japan.
- Ramey, C. H. (2006). An Inventory of Reported Characteristics for Home Computers, Robot, and Human Beings: Applications for Android Science and the Uncanny Valley.
- Rizzo, P. (2000). Realizing believable agents: An integration of the 'author-based' and the 'model-based' approaches. *AI Communications*, 13, 145-168.

- Reilly, W. S. N. (1995). *The Art of Creating Emotional and Robust Interactive Characters*. Paper presented at the AAAI Spring Symposium on Interactive Story Systems, Stanford University.
- Reilly, W. S. N., & Bates, J. (1992). *Building Emotional Agents* (CMU-CS-92-143). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Stein, B. E., & Meredith, M. E. (1993). *The merging of the senses*. Cambridge, Mass.: MIT Press.
- Sterling, R. (1959). Mirror Image (J. Brahm, Director), *The Twilight Zone*.
- Stoffregen, T. A., Bardy, B. G., Smart, L. J., & Pagulayan, R. J. (2003). On the nature and evaluation of fidelity in virtual environments. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual & Adaptive Environments: Applications, Implications, and Human Performance Issues* (pp. 111-128). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Thomas, F. & Johnston, O. (1984). *Disney: The illusion of life*. New York: Abbeville.
- Viviani, P. (1990). Motor-perceptual interactions: the evolution of an idea. In M. Piattelli Palmarini (Ed.), *Cognitive Sciences in Europe: Issues and trends* (pp. 11-39): Golem.
- Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception geometric illusions in dynamic displays. *Perception & Psychophysics*, *46*, 266-274.
- Viviani, P., Baud-Bovy, G., & Redolfi, M. (1997). Perceiving and tracking kinaesthetic stimuli: further evidence of motor-perceptual interactions. *Journal of experimental psychology: human perception & performance*, *23*, 1232-1252.
- Vinayagamoorthy, V., Steed, A., & Slater, M. (2005). *Building Characters: Lessons Drawn from Virtual Environments*. Paper presented at the Toward Social mechanisms of Android Science: A CogSci 2005 Workshop, 119-126.
- Welch, R. B., Warren, D. H. (1981). Immediate perceptual response to intersensory discordance. *Psychological bulletin*, *88*, 638-667.

Riferimenti

- Bryant, D. (2001). *Why are monster-movie zombies so horrifying and talking animals so fascinating?* Available: <http://www.arclight.net/~pdb/nonfiction/uncanny-valley.html>.
- Ferber, D. (2003). The man who mistook his girlfriend for a robot. *Popular science*.
oz project: <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/oz.html>
- robotjohnny. (2004). *Pixar and the uncanny valley*. Available: <http://www.robotjohnny.com/2004/10/04/pixar-and-the-uncanny-valley/>.
- Uncanny Valley, *Wikipedia*: http://en.wikipedia.org/wiki/Uncanny_Valley
- Uncanny Valley, *wordspy*: <http://www.wordspy.com/words/uncannyvalley.asp>
- Weschler, L. (2002). Why is this man smiling? *Wired*, *10.06*.
- Zoosis: <http://www.zoosis.com/corporate/n-index.html>

“Toward Social Mechanisms of Android Science. A CogSci-2005 Workshop”, [ICCS/CogSci-2006](#) Long Symposium (sessione speciale dedicata alla Valle del Perturbante), androidscience.com, http://www.benjamins.com/cgi-bin/t_bookview.cgi?bookid=IS%207%3A3.