

第一部 脳科学と人間観

人格の要件はカタログ化できるか

不気味の谷とパーソンネットワークの可能性

戸田聡一郎

1 はじめに

人格（パーソン）^①は、社会における意思決定において考慮されるべき重要な要素のうちの一つである。われわれは、他者の人格を否定するような行動や言動は取ってはいけないし、弱者の人格を尊重するように要請されている。しかも、われわれは何を人格とするべきなのかよく分かっている、つまり、人格を容易に直観しているように見える。しかし、「人格」そのものの定義を挙げる作業は、困難をきわめる。たとえば、ヒト胚は人格なのかという考察をしているときには、明らかに深遠な人格の条件を焦点にして議論しているように見える。他方で、コピー機の調子が悪く、私が機械に向かってうまく動くように言葉をかけながらコピーを取るとき、私はコピー機に人格を見いだしているのだ、と言うこともできよう。なぜわれわれはある存在者が人格であることを区別できるのか。そしてもし区別できるとするならば、人格とは果たしてどのような実体なのか。この大問題に答えるためには、哲学的・倫理的な議論を綿密に行うことがもちろん必要であろう。しかしその議論の手前において、より客観的な事実——つまり、区別できると思っていることがど

のような事実を指しているのかということ——を扱う必要がある。われわれは、「人格」をどのように表象しているのだろうか。その事実をめぐる議論は、哲学的議論に妥当性を持たせるさいの重要な前提となるはずである。そして人格の表象の解明は、引き続き次のような問題に連動している。その表象は、生命倫理学において、たとえばエンゲルハートが提唱したような、「社会的人格」（エンゲルハート一九八八）に意味を持たせることができるのか。それは今まで構築されてきた人格の要件とどのように関わることか。さらには、将来どのように人格の要件が構築されるべきかという問題に対して、何らかの解答を用意できるのか。

これらの問題群に見通しをつけるために、本稿では、人格概念に関わる脳神経科学的研究の成果と、それらが持つ倫理的含意について検討を行う。ここでの議論の基盤となるのは、近年の認知科学の知見と、二〇〇七年のファラーとハバーラインによる論考 (Farah and Heberlein 2007a) である。ファラーらは、人格を脳神経科学的に還元する可能性を論じ、その作業を通して人格を表象する「パーソンネットワーク」を提唱した。彼女たちによれば、そのネットワークはわれわれが生得的に持っているもので、他の認知機能とは独立に動員されるものであり、結局のところ、人格は錯覚である。そしてそのような錯覚は放棄すべきである。

こうしたラディカルな主張に反論を行うのは比較的容易であろう。しかし本稿では、ファラーらの主張の擁護を試みる。そのうえで、それでもなお彼女たちの議論には説明が尽くされていない部分があり、その不足部分に対する考察こそが、人格の議論の脳神経科学的解明にとって大きな意味を持つことを示したい。人格は直観によって語られるべきではない概念である。だがその概念の源泉となる、他者の心的状態を推測する機能と、その推測をもとにした他者の人格に関する認知は、たとえ否定されても直観として存在する。したがって概念としての人格の妥当性は、われわれが、その直観をコントロールし、乗り越える能力があるかどうかにかかっている。このことを有効に説明するためのヒントが、ロボット工学の「不気味の谷」仮説に

隠されている。

さらに、人格というデリケートな問題を扱うにさいして、フアラールらが人格を放棄すべきだと主張せざるをえなかった科学的・倫理的背景をも同時に考察しなければならぬだろう。

扱うべき事項は多いが、ここでは人格の要件について回答を用意するための予備的議論を行うことが最終目標である。そのことを確認して、まずはフアラールの議論を、適宜補足を加えながら、追うことにしたい。

2 フアラールの主張——人格は錯覚である

フアラールによれば、人格概念は、現象をうまく説明しているかのように見える錯覚である。人格なるものをわれわれの脳が表象するとき、それは現実を反映していない (Ibid.)。これが、彼女たちの結論である。なぜ人格概念の脳神経科学的基盤の解明を試みることで、このような一見ラディカルな主張に至るのだろうか。この論証の経緯を明らかにする必要がある。⁽²⁾

彼女たちはまず、「人格」の伝統的な定義を概観することから始める。グレイらのコメントリーが指摘するように、これらの定義は、必要条件を表すものであっても、必要十分条件を表しているとは言えない (Grey et al. 2007)。論文中で挙げられる哲学者たちは、ロックやカント、さらにこの論文によってフアラールの論敵となるであろうトゥーリーやエンゲルハートであるが、ここでは人格の定義の、ある意味で救いようなさを表しているともいえるフレッチャーの定義を取り上げよう。フレッチャーは人格を定義する一五もの条件を挙げ、そのなかでも知能指数 (IQ) が四〇以下であればおそらく人格ではなく、二〇以下では確実に人格ではないと述べる。これは明らかにわれわれの直観に反するものであり、客観的な指標を作ることの難しさを浮き彫りにしている。ほかの条件として彼が挙げるのは、自己意識や時間感覚、記憶などといった、

「ありふれた」条件である。しかし、そのようなありふれた条件を列挙してもやはり、客観的な基準を作るのは困難である。³⁾ 人格は、日常的に使われる便利な言葉であるにもかかわらず、その概念の満足な規定は不成功に終わっている。

このように、規定が困難であるにもかかわらず、人格は事実上、倫理の基礎概念となっている。ピーチャムとチルドレスによる生命医学倫理四原則が発達してきた背景にも、人格をいかに扱うべきかという問題があった。生命倫理でも議論されてきたのは、道徳的客体としての人格である。胎児は人格か、遷延性植物状態にある患者は人格か、という議論は、その胎児と患者の処遇についてわれわれに問いかけるものである。その概念は、「尊厳」という言葉と関連が深いものであった。そのような人格が生命倫理において問題になるのは、ある人（もの）が、ある人には人格であるように見え、別の人には非人格に見える事態が起こったときである。この事態において、人格を基礎においた生命医学倫理四原則を適用すべきかどうかという問いは不明瞭なものとなる。言い換えるならば、人格の定義の欠如が、問題の解決を困難にしていると言えよう。だが、そのような欠落を補うべく人格の基準を正確に列挙し、人格がどのような構成によって成り立っているのかを明らかにするのは、予想以上に（あるいは予想通り）難しい。問題は、人格を構成するとわれわれが直観している心理学的特徴を、特異的な定式化に落とし込もうとすると、われわれの直観が置き去りにされるといふことである。人格が成立するために、どのような特性のセットを持つべきなのか、そしてその特性をどの程度持つべきなのかは定かではない。その条件は、先ほどのフレッチャーの定義のように、任意に決めることが可能であると言っていいただろう。

概念的・原則的な議論の試みが失敗したあとで考えられる方法は、人格を再び経験論の俎上に乗せることである。経験的議論において現在有力なのは、脳神経科学の知見であろう。もし人格の特徴を實現化しているものが大脳皮質の働きであれば、それらの特徴を表現する皮質活動の基盤を探ることができる。遷延性植

物状態患者は、生物学的には生きてはいるが、大脳皮質の機能を失っており、心的生活を失っているように見えるので、「かつて人格であった人」とみなされることが多い^③。もし皮質活動が人格概念の構成にとって重要であるならば、膨大な認知神経科学の知見が適用可能だろう。認知神経科学は、人格の条件である合理性や知能を構成する要素、たとえばワーキングメモリや自己モニタリングなどの心理学的特徴や、将来の計画や言語でのコミュニケーション機能を、ある脳領域群の活動という「脳神経科学の言葉」に置き換えてきた。しかしながら、そのような探求が十分になされたとしても、依然として、人格概念には解決すべき問題が残っている。つまり、客体のどのシステムが、どの程度働けば、人格の基準を満たすことになるのだろうか^⑤。

以上を考えれば、現在の脳神経科学が貢献できるとするならば、それは、何が人格なのかという問題の探求においてではなく、なぜわれわれは「人格である」という直観を持つてしまうのか、という問題の探求においてである。この直観は、フアララーの考察によれば、人格と非人格を区別してきた経験から生ずるものではない。それは生得的で、初めからわれわれの経験を構成しているものである。この主張が最終的な（ニヒリスティックな）結論につながる。すなわち、われわれは人格、非人格の区別を明確に行っているように見えるにもかかわらず、その判断を基礎づける脳内のネットワークは外界の本性を客観的に反映しているものではなく、よって人格は錯覚なのだ、という結論だ。この結論が本稿における検討の対象となる。

この結論に至るにあたってフアララーは、人格の表象機能を、大脳の活動ネットワークに求める作業を行う^⑥。もし人格が物理的性質を表象する概念であれば、それに対応した脳の表象があるはずである。フアララーによれば、その物理的性質の表象は、「パーソンネットワーク」なる神経回路に強く関連している。このネットワークは、紡錘状回、扁桃体、側頭ー頭頂接合部（TPJ）、前頭前野内側部という四つの領域からなる。ここでフアララーが「パーソン」として意図しているのは、人格の存在の直観の源泉となり、かつ単

純化できる特性であることに注意しておこう (Farah and Heberlein 2007b)。紡錘状回は、ヒトの顔に特異的に反応する領域である。⁽⁷⁾つまりここでは、「ヒトの顔」が人格の源泉の一つとして想定されている。また扁桃体は、表情認知（とくに恐怖反応）に関連した反応を見せ、感情にかかわる原始的な機能を担っている。⁽⁸⁾われわれは、提示される刺激が顔そのものでなくても、目の表情があるだけで、そこから得られる人格の感情をくみ取ることが可能である。TPPJは、他者の目的を持った行動に対して特異的に反応を見せる部位である。たとえば、「ポイント・ライト・ウォーカー（またはバイオロジカル・モーション）」、すなわち多くの光の点だけでヒトの動きを再現している画像は、ランダムに動いている光の点（スクランブル・モーション）と比較して、TPPJの特異的な活動を引き起こす。⁽⁹⁾最後に挙げられる内側前頭前野は、他者の心の評価を行うときに、さまざまな文脈において活動を見せる。人物が登場する物語を提示されたとき、その物語を理解するさいには内側前頭前野が関与しているという研究が、多数報告されている。物語に人物が登場しない場合は、この領域の活動は見られない。さらに人物の印象を評価するとき、またじゃんけんゲームにおいて、コンピュータではない他者の次の手を予測するとき、課題に対する内側前頭前野の、より相関の高い活動が得られている。⁽¹⁰⁾ここでは、最終的には、他者の心的状態の推測が重要な役割を持つ「社会脳 (social brain)」に繋がるようなネットワークが提示されているといえよう。

フアラールはこのようなネットワークを論じたうえで、重要な示唆を行う。すなわち、これらネットワークは自動性と生得性を持っており、人格と非人格の判断はこれに対応していない、というものである。

ここでフアラールがいう「自動性」とは、たとえ刺激が人格でないと分かっているにもかかわらず、ある刺激に対してはパーソンネットワークが動員される、という意味で用いられる。スマイリーマークや、棒で構成される人間の図形に対しても、パーソンネットワークの活動が引き起こされる。また、複数の図形が追いかける場合、アニメーションを見たときには、われわれは容易に図形の「意図や欲求」をくみ取ることができるだろう。そ

のような動きに一貫性を有する図形群に対しては、機械的なランダムな動きを知覚しているときと比較して、パソコンネットワークが活動する。パソコンネットワークの一部である扁桃体を失った患者は、図形の物理的特徴しか報告できないことが知られている。また、健常被験者が独裁者ゲーム（ある金額を他者と分け合うさい、その分け前を被験者自身で決定することが可能であり、その決定は必ず実行されるゲーム）を行っているさい、コンピュータのデスクトップ上に、目と鼻だけの顔の絵が描かれているだけで、被験者はより寛大な戦略（相手に、より多くの分け前を与える戦略）を採用したことが報告されている。この結果は、人格を認知していなくても、パソコンネットワークが自動的に動員され、その活動が行動に影響を与えるという強力な証拠となるだろう。

他方で、ファラーらのいう「生得性」とは、「人格が存在する」ということを学習しないでも、人格性を読み取るシステムが遺伝的にあらかじめプログラムされている、という意味で用いられる。ここでは乳児・幼児の研究が有効になるだろう。紡錘状回を早期に損傷した男の子の症例では、彼は成長して他の物体を認識することはできるようになったが、ヒトの顔は認識できるようにはならなかった¹⁾。また、さきほどの図形群のアニメーションについても、いくつかの示唆的なデータがある。生後三か月の乳児は、「意図」を持つ図形群のアニメーションに選択注視を見せる。さらに、自閉症児のデータも有効である。自閉症は、他者の心を読む「心の理論」が欠如しているという仮説のもとに、その遺伝的要因が探求されている疾患である。他者の心が読めないことが主な原因となつて、自閉症児は、生まれたときから、人格性に関連するような意図を持つものを、単なる物体のように扱う。その行動を裏付けるように、自閉症児にヒトの顔を提示しても、紡錘状回は活動せず、さきのアニメーションの動きに対するパソコンネットワークの活動は認められない。

このようにパソコンネットワークの自動性・生得性を指摘したうえで、ファラーらは、人格概念が、現象をうまく説明しているかのように見える誤った形而上学的概念である、と結論する。この世界には、明らか

にわれわれと同じような心を持っているものがあるし、そうでないものもある。おそらく、心にはさまざまな種類・程度があるのだろう。しかし、そのさまざまな心に対する評価は、表面的な、ミスリーディングかも知れない手がかりによって決められている部分がある。したがって、人格の表象は、現実を正確に反映していない。この結論をもとにして、最後にフアラールは、倫理的問題を解決するにあたってわれわれに残された唯一の選択肢は、人格概念を廃して、功利主義的なアプローチ法をとることだと述べる。

ここではフアラールの議論を、おもに脳神経科学的アプローチの可能性について焦点を当てながら追うことを試みた。彼女たちは、人格の基準の实在可能性について述べたあとで、その基準に依拠しないような人格の直観の仕方を、脳神経科学の知見に求めた。その結果、われわれの人格の直観の仕方は、人格の「概念」を認識する方法とは異なることが判明した。人格の直観は、たとえその対象が非人格と分かっているにもかかわらず、視覚における錯覚と同じような仕方である。したがって、混乱した倫理的帰結を生むのである。

3 フアラールの主張の位置づけ——人格は放棄されるべきだったのか？

前節において、フアラールが展開した人格の定義のあいまいさの指摘、および「パーソンネットワーク」の提唱を概観した。しかしながら、これら個々の指摘・提唱については、新規性があるわけではない。このことをまず確認する必要がある。パーソンネットワークは、「心の理論」を基礎づけるものとして、すでに以前から提唱されていた（たとえば Allison et al. 2000; Gallagher and Frith 2003）。われわれが使用している語の定義の境界のあいまいさも、フアラールに対するコメントや他の文献が指摘しているとおりである（たとえば Banja 2007; Churchland 2007; 信原二〇〇九）。しかしながら、フアラールは、既存の知見から、倫理学

の基礎となるべき人格概念の錯覚性を指摘し、それを放棄すべきことを論じたのである。本節では、フアラールの議論のどこが重要性を持つかを検討する。議論の力点を探ることによって、解決すべき問題点はどこであるかを明示したい。そのために、フアラールが結論を導出した政治的意図と生命倫理学的背景についても考察する必要がある。一見、遠回りする議論ではあるが、ここにはフアラールがおそらく意図的に見落としている重要な示唆が含まれている。

人格の定義のあいまいさについては、すでに述べたように、新規性があるようには見えない。しかしながら人格概念と、人格以外の概念との間には、ある重要な違いがある。一般に、人格以外の概念においては、山と丘（標高）、禿げと禿げでない（頭髪の量）、といったように、概念を切り分けるべき境界線があいまいであっても、明確な基準・変数（標高と頭髪の量）が存在する。だが他方、人格概念は、その基準・変数すら持たないにもかかわらず、われわれに人格をはつきりと認識させる力を持つ。そしてその基準を新たに作るうとすれば、フアラールが最初に力説したように、形而上学的概念や社会的経験に頼らざるを得ず、それらの試みは失敗に終わるのである。このような、「基準が」フィクションであるのにリアルに感じ取ってしまう「人格の特性が、まさに錯覚である。ここでのフアラールの錯覚にかんする議論は、二重の意味で用いられることに注意したい。一つ目は、人格と非人格を区別する基準は実在しているようで実在していないという意味である。二つ目は、われわれは、いくら努力して対象が非人格であることを自分に言い聞かせても（たとえば図形同士の追いかけ合いのように）適当な刺激さえあれば、そこにリアリスティックな人格を自動的に感じ取ってしまうという意味である。

しかしここでの「リアルに感じ取る」特性を持つ人格は、倫理規範としての人格と比して、明らかに性質が異なるものだ。フアラールがいう「パーソンネットワーク」とは、この「リアルに感じ取る人格」を指している。それは決して概念ではない。乳児が図形たちの追いかけ合いを愛好注視するとき、その乳児は人間

らしい刺激を「感じ取り」、その刺激をより注視するが、「この図形は人に値する諸条件を持つ人格だ」と理解しているわけではない。リアルに感じ取る特性を持つ人格、すなわち「パーソン」は、たしかに、「人格」概念の源泉のひとつではある。そして先述したように、「パーソン」が二重の意味で錯覚であるということは、論証可能だろう。ファラーはこの意味での「パーソン」を「人格」と仮定して、その概念の無益性を論証したのであった。しかしその議論は、人格「概念」が錯覚であり、その概念を放棄すべきだという結論には結び付かない。パーソンの錯覚性の議論は、社会規範としての人格概念の是非を問うには弱すぎる。そして、人格概念における基準の不在は、わざわざパーソンネットワークを持ちだして脳神経科学的事実を列挙しないまでも、十分に指摘することができるのである。

ではなぜ、ファラーらはパーソンネットワークという弱い証拠から、人格概念を否定しなければならなかったのか。そこには、生命倫理学の基礎をなしてきた人格論（パーソン論）に対する経験的評価を行おうという試みが見てとれる。ここでいう経験的評価とはもちろん、脳神経科学の知見による検討である。もともと、人格論における人格と非人格の区別は、日常生活に根差したものであった。それは、解答を出すうえでの方便であるとさえ言える（香川二〇〇九）。その議論の社会的影響について、金森（二〇〇七）は、生き生きと記述している。金森によれば、

生命倫理学という学問の中で提示されるパーソン論^⑩は、理性を顕揚するということ自体よりも、〈理性的存在〉から見て、亜・存在や準・存在に対する処遇に、ほぼ一貫して厳しい評価を下すことに大きなエネルギーを費やしてきた。「……」ごく素朴に見ても、それが或る種の冷淡さや冷酷さを抱えていたというの、否定しがたいように思える。「……」興味深いことに、本国のアメリカでも、この学問自体が、よってたかつて人々を〈死への傾斜〉に傾かせるとして、必死の批判をしているスミスのような人もいる。

「……」こうして、生命倫理学の中でのパーソン論は、障害者差別や弱者排除を正当化する危険な議論としての位置づけを受け、それに対しては一定程度の留保や距離設定をすることが共通の了解事項のような様相さえ呈しているといっても過言ではないのである。(同、一一四～一六)

このように、人格概念は、われわれの日常生活における判断を基礎においているにもかかわらず、評判が悪い。その議論の妥当性を検討するには、われわれの日常生活の判断の本性を調べる必要があるであろう。そしてそのためにファラーらが持ちうる方策が、脳神経科学であった。彼女たちはコメンタリーへの応答のなかで、人格性をめぐる問題の調停に脳神経科学が独自の解決策を提示するわけではなく、人格性の自然化ができるとは考えられないものの、道徳理論と自然界 (natural world) の関連性を調べることには意味があると主張している。彼女たちによれば、「道徳原則はそれ自体、経験的評価を必要とはしないだろうが、原則とは、現実世界の実体に言及しているのであり、とくに生命倫理学では、経験的な現実においてそのような原則をどのように支えるかが重要となる」(Farah and Heberlein 2007b, W3)。人格概念が生まれた経緯を見れば、それが自然化されるとはとうてい考えられない。コメンタリーを寄せたペリングも、人格は科学の概念ではないことを強調する (Pering 2007)。「解答のための方便」は科学的に説明できないにもかかわらず、それでも自然界と何らかの関係を持つていないか——それがファラーらの(最初の動機ではないにせよ)意図するところであった。しかし、その意図を確認したうえで、もう一度ファラーらの論証を見てみると、やはり、パーソンが錯覚だから人格概念は信頼すべきでないという議論の不自然さが浮き彫りとなる。それは経験的事実を規範的概念に組み込むという自然主義的誤謬を犯している (Meyers 2007)。そしてそのやり方は、かなり強引である。ファラーらのいう経験的事実とは、他者の表情認知や他者の心的状態の認知にかかわる、いわゆる「心の理論」に関連した脳内のネットワークであった。それがパーソンネットワークと銘打たれ、

そのネットワークの自動性・錯覚性から人格概念の危うさが導かれる。そうして彼女たちは人格概念を抜きにした功利主義的アプローチを提唱するに至る。フアラールは、フレッチャーの人格概念の条件を借りれば、多数の条件のうちの一つ、「他者とかかわりが持てること」の能力について、脳神経科学的知見を援用して自然化を試みたにすぎない。そしてその知見は、他者とかかわっている本人が「どのように、何を考えながら」かかわりを持っているかについて、きわめて限定的な情報しか提供しない。これは、脳神経科学のアプローチ方法の限界を考えれば、当然のことといえる。しかし、まさにこのかかわりのあり方こそが、人格概念にとって重要なのである。したがって、フアラールが示す科学的事実、人格概念まで手が届いておらず、その概念の是非について議論できるまでの証拠群ではない。彼女たちは、人格論が「生命倫理学にとって距離をとるべき」学説だということを確認し、それを科学的に示したかったのかもしれない。だがそのことは今のところ主張できないだろう。

しかしながら、彼女たちの試みは、決して無駄ではない。本稿ではフアラールの意図を最大限擁護したい。その意図とは、人格概念は、自然界に繋がっており、科学的アプローチが可能である、というものである。彼女たちはその繋がりをわずかに示しながら、やはり両者の調停はできないのだと結論している。もう一度フアラールの分析の仕方を確認しよう。まず、道徳的客体を対象にし、人格概念について脳神経科学的に解明しようとする試みは、成功しないだろう。なぜなら、人格の基準が存在しないために、どの脳領域が、どの程度活動すれば、その客体が人格であると判別すべきかがわからないからである。そこでフアラールは、道徳的主体に関心を移す。われわれはなぜ、そしてどのようにに人格を感じ取るのか。客体の機能を感じる主体の脳という、脳神経科学が採用してきたアプローチまで、人格概念を落とし込んだところに、フアラールの独自性があった。その結果キーワードとなったのが、パーソンネットワークだった。しかしながら、あるいは当然のことながら、フアラールの論考においては、道徳的主体のもう一つの脳機能には注目していな

い。それは客体の機能を感じ取り、その知覚を主体として統合し、人格かどうかを判断する機能である。つまり、人格概念の自然化を試みるには、客体の人間的な意図を読み取るという機能（ファラーらのいうパーソンネットワーク）に加えて、それを吟味し、「パーソン」ではなく「人格」として受け入れる機能が付加されなければならない。

もちろん、この付加されるべき機能は、道徳的主体の脳機能のなかでも、どの領域がどの程度活動するかという問題を扱わなければならない。ファラーらが検証不可能だとするネットワークと同じ欠点を抱えているかもしれない。しかし、ファラーらが言う問題点とは、胎児や遷延性植物状態患者のような客体の脳内表現をわれわれが見るときに生じる問題である。客体（胎児や意識障害患者）の心的状態について、われわれが彼らの脳内表現を手がかりに推論したうえで、その推論が人格の基準を満たしているかをさらに推論する——この作業は仮定に仮定を適用することであり、信頼性が乏しい。対して、人格性を判断している主体の脳を見ることは、主体の人格性に関する判断という明確なアウトプットと関連させることができる。さきの金森の言葉を借りれば、〈理性的存在者〉から見た客体（垂・存在や準・存在）についての判断を、主体の意思決定・行動と関連させながらみていくことは可能であり、人格概念の自然化には、概念の本性を考えれば、その作業は不可欠である。次節以降では、この主体の概念的判断をめぐるネットワークの可能性について、論証を行う。そこでカギとなるのが、ロボット工学における「不気味の谷」の仮説である。

4 パーソンネットワークに付加されるべきネットワーク——「不気味の谷」を手がかりに

客体のパーソン性を見いだすことと、主体がそれを判断すること、この二つの機能の統合が人格概念の自然化には欠かせないことを述べた。後者をより明確に示すためには、パーソンネットワークの証拠を提供し

た実験の環境に比して、より相互作用可能な、かつ、人格性にかんしてより反省的吟味が必要な状況が用意されなければならぬ。そしてそれは、たとえば胎児をどのように扱うべきかというような、長い議論の伝統を持ち、先入観が混入しがちな問題よりも、より新規性を持つトピックのほうが望ましい。主体（人格の判断をする者）と客体（準・存在や亜・存在）が相互作用することで、パーソン性、人格性が変わるような状況があるだろうか。そのヒントが、こんにち長足の進歩を遂げているロボット工学に潜んでいる。「不気味の谷」の現象はまさに、パーソン性・人格性が複雑に絡んだ性格を持ち、主体と客体の時系列的な相互作用が可能な状況を生み出す。

「不気味の谷 (uncanny valley)」は、一九七〇年、ロボット工学者の森政弘が提唱した仮説である（森一九七〇）。彼は、ヒト型のロボット（アンドロイド）を人間に似せるようにしながら開発していくときに人間がアンドロイドに感じる親近感に着目した。横軸に、アンドロイドがどれだけ人間に似ているかを計測した尺度を、縦軸に、人間が感じる親近感をとり、グラフを書く（図1）。アンドロイドが人間に全く似ていない場合、最初はその動作や外観が人間に近づくにつれて、親近感は増大していく。しかし、人間に近づき過ぎたある時点において、その地点以降、アンドロイドの不自然さが急激に際立つようになる。そして親近感は反転し、ネガティブな感情を引き起こす。その感情は、不気味なもの (uncanny) であり、ぞつとするもの (creic) であり、嫌悪感を引き起こすもの (disgusting) である。親近感が最小値をとるときには、人間はそのアンドロイドを「まるで生きている死体」であるかのように感じるであろう。だがもし、アンドロイドが人間に限りなく近づいて自然なコミュニケーションがとれるならば、そのネガティブな感情は再び急激に反転し、親近感を抱かせるはずである。この一連の感情の流れ、図1における親近感の急激な落ち込みと回復が、「不気味の谷」と解釈される。そして、その不気味さの効果は、動かないアンドロイドより、動くアンドロイドにおいてより一層増幅されることが予測される。以上が、森の提唱した不気味の谷の仮説の

概要である。

この仮説は、一種の予言であり、アンドロイドを製作する技術の限界のために、科学的検証がなされてこなかった。しかし、近年になって、不気味の谷現象の存在が、静的な画像を用いた研究であるが、明らかになっている (MacDorman and Ishiguro 2006; Seyama and Nagayama 2007; MacDorman et al. 2009)。また、サルにおいても、不気味の谷が再現できることが報告されている (Steckenfinger and Ghazanfar 2009)。そして、不気味の谷の左側（人間にまだ「似すぎていない」側）と右側（それを実際の人間と同一視した場合）にかんしては、より実証的な研究が行われている (Gong 2008; Kachi et al. 2008)。

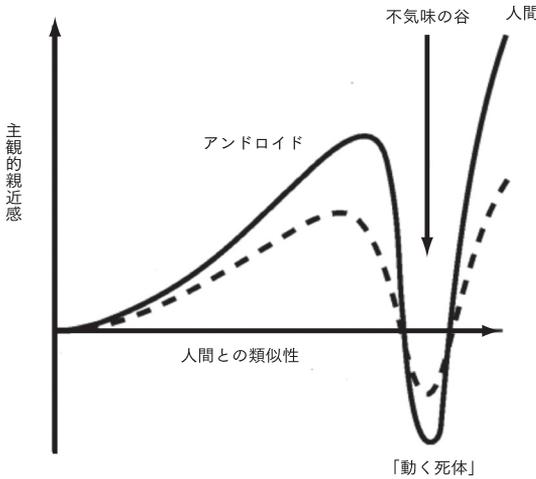


図1 不気味の谷。実線は動くロボットに対する親近感、破線は静止しているロボットに対する親近感を示す (Steckenfinger and Ghazanfar 2009 をもとに改変)。

クラックらは、被験者にゲーム（囚人のジレンマゲーム）を行わせるさい、その時の対戦相手をコンピュータ、アームロボット、人間に近い（とはいえリアルスティックな人間ではない）頭部を持った機械、そして実際の人間に振り分け、各条件での脳活動を計測した。最初に対面式でゲームを行ったのち、機能的核磁気共鳴画像（fMRI）で脳活動を撮像した。その結果、機械に心を帰属させていると考えられるときには、その外見が人間に近くなるほど、心の理論に関連する領域が活動を見せた。それらは、側頭―頭頂接合部（TPJ）や、内側前頭回であり、ファラーらが提唱しているパーソンネットワークと重なる部分が多い。注目すべきは、人間に近い頭部を持った機械に対して、被験者は、対面式でゲームを行っているさい、話しかけたり、文句を言ったりしたことである。われわれは機械にも心を見いだしているのか？心の理論において、どの領域が必要で十分なものは分かっていないが、答えはおそらくイエスである。そしてそれは、ファラーらが主張するように、自動的に起こる反応であろう。ゲームを行い、相手の有意な反応が「返ってくる」のであれば、なおさらである。しかしその反応は、不気味の谷の左側でしか検証されていない。人間の外観・動作には似ていないけれども、有意味な行動をとり、反応を示す刺激に対するわれわれの反応である。ファラーらは、多くの論者が疑問を唱えたパーソンネットワークの錯覚性を擁護するために、マサチューセッツ工科大学が開発したロボット、キスマット（Kismet）のビデオを見てほしいと反論した（Farah and Heberlein 2007b, W2）。あるいは、日本の国際電気通信基礎技術研究所（ATR）が開発したロボビー（Robovie）でもよいだろう。確かにわれわれは、これらロボットの動画を見ると、それらを単なる機械ではなく、無条件に、意図を持つ行為者として見るであろう。だが神田らは、ロボビーの製作にあたっては、不気味の谷を避けるために、意図的に「機械的な声」「機械的な動作音」を出すように設計したことを指摘する（Kanda et al. 2004）。キスマットもその外観からして同様に、不気味の谷を避けることを意図して製作されているだろう。無条件にロボビーやキスマットにパーソンを感じてしまうのは、その外観と

機械音のおかげだからとも言えよう。したがってフアララーのいうパーソンは、不気味の谷の手前でしか有効でない。パーソンネットワークはたしかに自動性があり、そのおかげで、単純な刺激に対する反応の説明が容易になっている。しかし、たとえば人間に近いが人間というには明らかに不自然であるような、複雑な情報処理を必要とする刺激に対しては、フアララーのパーソン性は無力になるだろう。不気味の谷においては、パーソンネットワークに加えて、それとは異なるネットワークが動員されている。そして、本稿が主眼に置くのは、それらネットワークを超えて理性的に判断するネットワークが存在するかどうかということである。不気味の谷の底にいるアンドロイドに対してわれわれは、拒絶感を示すかもしれない。それは人格の判断を混乱させるものである。不気味さと人格の判断は、同一のものではないが、互いに干渉し合う。それでもわれわれは〈理性的存在〉として機能し、それらを道徳的共同体へと迎え入れることができるのか。あるいはやはり、パーソンネットワークを基礎に置いた自動的な脳機能にわれわれは身をゆだねるしかないか。

あらかじめ断っておけば、この問題を解決できるだけの科学的証拠はまだそろっていない。しかしその問題の解決がはたして可能なかどうかは、検討する意味がある。その解決可能性を説得的に示すためにも、見ておくべき事実はまだ存在する。

5 人格性の自然化の再挑戦——倫理学の人格論を救い出せるか？

前節で述べたように、不気味の谷はパーソンネットワーク以外の他の説明を必要とする。

では一体どのようなネットワークが、この「不気味さ」を構成しているのだろうか。ここでいう「不気味」とは、フロイトの記述にしたがえば、次のようになるだろう。「不気味さとは、外部から来た異星人や見知

らぬ何かに由来する恐怖ではない——それとは反対に、われわれから切り離そうといくら努力してもそれに対抗できないような、奇妙に親密性をもつ何かに由来する恐怖なのである」(Fried 2003, 135-4)。不気味の谷の説明には、いくつかの説が存在する。有力視されているもののひとつは、期待に対する裏切りである (MacDorman and Ishiguro 2006; Seckenfinger and Ghanzafar 2009)。われわれは、限りなく人間に近いアンドロイドを見せられると、それが人間らしく振舞うように期待するだろう。しかし、実際のアンドロイドの運動は、ぎこちなく、不自然である。この認知がネガティブな感情を引き起こし、われわれは不気味の谷に「落ちる」ことになる。しかし、期待の裏切りに関連したネガティブな感情がなぜ「不気味」という特異な感情に結び付けられるかについては、まだ説明されていない。一般的に言って、刺激に対する期待が裏切られただけでは、不気味さは発生しないだろう。また、この現象にはいくつかの進化的な説明が付けられている。たとえば、病者の探知や、死者への忌避感、生殖の有利さに基づいた説明など、目的論的な議論が展開されている (MacDorman and Ishiguro 2006)。先述したような、サルにおいても谷の存在が示唆された研究結果は、進化的説明を補足できるかもしれない。だがどのように十分な進化的な説明を加えても、それらは不気味さを構成する対象が持つ性質や、不気味さを感じ取るわれわれの認知プロセスの解明には容易に繋がらない。

しかしながら、ロボット工学においては、不気味の谷が回避されれば、そこで問題は解決される。現在のロボット技術は、よくメディアに登場するように、もはや不気味さを感じさせないまでに精巧に作り上げられている。また、仮想空間において、アバターとしての人体を再現できれば、容易に谷を「飛び越え」、それが社会的な脳機能の解明に大きく貢献するかもしれない (Moser et al. 2007; 藤井二〇〇九、一九七〜九八)。現在の技術がこのまま進めば、アンドロイドを含めたロボットは、容易に谷を飛び越え、社会の一員としてわれわれの共同体に組み込まれるかもしれない。アンドロイドがわれわれの「仲間」として迎えられるよう

な社会を想像することは、もはや夢物語ではない。このように、不気味の谷は回避可能であるから、その神経基盤の研究の重要性は、以前よりも薄れているかもしれない。

だが本稿の主眼は、谷を回避できる実現可能性とはあまり関連性を持たない。谷を最初から回避するのではなく、あえて、不気味の谷へと落ち込む現象を想定してみよう。ここでは、アンドロイド（またはアバター）という準・存在、亜・存在的なオブジェクト自体が技術改変されるわけではない。同一のアンドロイドに対して、コミュニケーションをとり相互作用を行うことで、われわれが何らかの判断を下して彼（女）を仲間として受け入れ、自然に対応することができかねるかが検証される。この時系列的な機能の変化は、不気味の谷現象においてまだ検証されていない。われわれは、意図的に、不気味の谷を「乗り越えられる」だろうか。谷を跳躍するのではなく、意図的に乗り越えること——ここに理性的存在者の入り込む余地があり、人格性のもう一つの源泉がある。不気味の谷の谷底にいたとしても、アンドロイドの外観はある程度人間に近く、その意図は汲めるはずである。したがってファラーの主張するパーソンネットワークは確実に機能しているだろう。もちろん、不気味さを乗り越えることと、アンドロイドに対してきちんと対応することは別の機能である可能性も十分考えられる。しかし、その対応の仕方と、不気味さとの関係を追うことは、人格のボーダーラインの事例における心理学的機構の解明に寄与するものであり、そのどちらもパーソンネットワークを超えたシステムである。

これらのコミュニケーション変化は、われわれの意識的な思考の末に決断されるものかもしれない。反対に、ファラーらが指摘したように、パーソンネットワークによって自動的に処理されて引き起こされるものかもしれない。不気味の谷が、心の理論に関わるシステムや、ミラーニューロンシステムのみで解決される可能性は、依然として棄却されていないのである。かくして、不気味の谷からの回復過程を探ることによって、人格性の判断とはどのような意思決定なのか、明らかとなるであろう。われわれは、ボーダーライン

にいる客体について、やはり自動的な脳機能のみによって人格性を認め、それを共同体の中に迎え入れるのか。それともやはりわれわれは〈理性的存在〉であり、その理性こそが、共同体への受け入れには重要であり、それが人格概念を基礎づけることになるのか。そしてもし、アンドロイドについて社会的人格（エンゲルハート一九八八）という概念が適用されるならば、それは功利主義という蓑をまとった錯覚なのか、それとも自動的処理に惑わされることなく理性的に判断された結果なのか。先述したように、そのすべてが明らかになるとは考えられないし、人格性の判断も、理性か感情かというような二者択一なものであるとは考えられない。しかし不気味の谷を乗り越えていく過程を追うことで、人格の判断と、その判断に必要な要件についての重要な示唆が得られるに違いない。もし、コミュニケーションが取られているにもかかわらず、彼（女）らを拒否し続けるならば、やはりわれわれは、自動的な処理に強く縛られていることになるだろう。

フアララーらの議論と、不気味の谷仮説を追うことで、パーソンネットワークでは扱いきれないような事例があることを確認した。実際、フアララーらの議論は、倫理的問題におけるボーダーラインを問う事例について、人格概念が無力であることを示そうとしたものである（永守二〇〇七、一六二）。本稿ではあえて、新規性のあるボーダーラインの事例を問うことで、人格概念と自然界との「細いつながり」を自然化する可能性を検証した。実際のところ、人格性をめぐる判断は、不気味の谷においても、胎児の人格性の問題と同様に、多様性のあるものかもしれない。もつと基礎的な脳機能である共感性について言えば、その多様性は広く認められているし、一個人内でも状況によって判断は異なる（Keyser and Gazzola 2007）。われわれは新しい倫理的ボーダーラインにおいて、その脳機能を追うことによって、より新しい人格の基準が見いだすことができるか。これまで議論してきたように、この問題に対する解答はまだ用意できない。さしあたりは、コミュニケーションや学習による人格概念の変容の可能性が指摘できるところとどまるだろう。加えて、認知過程を追うための方法論についても、大きな壁が立ちはだかっている。被験者にどのような質問を行えばよいの

かも、今のところ明確でない。評定法 (rating) のような、被験者への単純な質問では解明できない部分も出てくるだろう。それでもなお、その方法論を含めた検討は十分な新規性があるし、何よりもこの問題系は、人格概念を基礎づける合理性まで射程に入れている。準・存在である他者とのかわりのあり方を記述することによってパーソンネットワークに加えられるべき認知機能は、人格の要件のカタログ化に向けて、重要なヒントとなるだろう。

自然界とパーソンと人格——この三つをつなげる科学的事実を積み上げていくことは、広く道徳の本性、そして将来には倫理の脳神経科学に関わる、重要な問題系である。

註

- (1) なお、本稿では、「人格」と「パーソン」という交換可能に見える語について、特段の断りがない限り「人格」を用いることとする。ただしフアララのいう「パーソンネットワーク」では、のちに述べるように、規範的な人格概念とは異なる文脈で「人格」を表現しているので、差異を強調するために「パーソン」ネットワークとした。本稿の後半部分、とくに第3節以降では、「人格」とフアララの「パーソン」はかなり異なる意味合いを持つようになる。
- (2) このフアラールとハバリーラインによる論文、およびその論文に対するコメントリー群について、永守(二〇〇七)がレビューとして的確にまとめており、この議論の意義も確認している。彼は「人間の耳を持つマウスのような」キメラ」に対する考察が人格の実践的使用に効用となることを示唆している。彼の示唆については、本稿を通して、別の視点から具体的な考察を行う。
- (3) フレッチャーの指摘した「ありふれた」条件のなかで、本稿にとくに関係があると思われるものは、「他者と関わりが持てる能力」と、「合理性と感情のバランス」である。いずれも素通りしやすい条件であるが、フアラールの「パーソンネットワーク」と、その後論じる議論において、重要な観点となる。なお、もともとフレッチャーは人格性 (personhood) ではなく、人間性 (humanhood) の条件としてこれらの条件を挙げた (Fletcher 1979)。
- (4) 遷延性植物状態患者の意識については近年議論が活発になってきた(総説としてOwen and Coleman 2008)。そしてそれらの議論の基礎をなす、患者の意識に必要なのは高次大脳皮質領域の有意義な活動であるという考え方は、より高次の認知機能に関連した皮質活動が人格にとって重要である、という前提に基づいている。意識障害におけるこの前提の是非はここでは論じない。
- (5) たとえば、大脳の発生に問題があり、生まれつき大脳を持たない無脳症児の意識の研究は、この問題に対して示唆的であろう。大

- 脳皮質の活動は存在しないにもかかわらず、無脳症児は兄弟に対して愛着を見せ、笑顔を見せる (Mether 2007)。そのような子ども
 笑顔を見ると、われわれはその子には人格があると考える傾向があるだろう。このときの判断を脳神経科学的に解明できるか。この
 問題は引き続きフアラールの提案を考えるさいの有効なトピックとなる。
- (6) 以降本節では、フアラールの取り上げている多くの脳神経科学の知見については、補足を加えつつも、文末の参考文献には挙げな
 いこととする。彼女たちと重複しない参考文献のみ、リストに加えた。
- (7) 紡錘状回の顔領域 (FFA) における機能特異性・領域特異性については、カンウィッシャーとヨーヴィルのレビュー (Kanwisher
 and Vovel 2006) を参照せよ。
- (8) 扁桃体は、刺激の内容が意識に上らずとも、表情を認識することが可能であり (たとえば Morris et al. 2001) 、このことは扁桃体の機
 能の自動性を表しているといえる。
- (9) また、このときの活動は、ヒトの体部位よりもむしろ、ヒトの動きそのものに対して特異的であることがわかっている (Peuskens et
 al. 2005)。
- (10) しかしながら、ここで言及されている機能は、のちにフアラールらが主張するネットワークの生得性・自動性を必ずしも支持してい
 ない。内側前頭前野を脳卒中によって後天的に損傷した患者は、記憶や計画の機能に障害を見せ、作話をする傾向にあったが、他者
 の心的状態の推測に障害をきたしてはいなかった (Bird et al. 2004)。
- (11) また近年、生後一定期間「顔」を見ないまま人工的に飼育されたサルが、初めて顔を見る期間を与えられると、(生まれてから一度
 も目にしていない) 顔に対して選択注視を見せたことが報告されている (Sigara 2008)。この結果は、顔に反応するネットワークが生
 得的であることを支持するものである。
- (12) この引用での「パーソン論」は、本稿における人格とパーソンの区別における「人格」を指す。
- (13) サルにおける不気味の谷の再現は、以下のような行動学的データの報告であった。被験者のサルが、同種の、真に迫るリアルな C
 G を見せられた場合には、実際の同種のサルの動画と、リアルでないロボット的な同種のサルの CG を見せられた場合と比較して、
 動画を注視する回数と時間が減少したのである。もちろんこのデータだけでは、サルの心的状態を正確に推測できているとは言いが
 たい。引き続き行われるであろう生心理学的指標の解析と、電気生理学的数据が、この現象の実証にむけた重要な研究となる
 はずである。
- (14) 各条件でのゲームに対する動機付け、および戦略は変わらなかったことが、行動学的データによって示されている。
- (15) またここでは、われわれがどのように「人間」の定義をしてきたか、また人種差別がどのようにして擁護できないものになったの
 かについての、文化的・歴史的考察 (Furukawa-Anusko 2004) も役に立つかもしれない。彼によれば人間の条件は「必要な神話」であ
 る。本稿での人格についてもそのような言説を無視するわけにはいかない。

参考文献

- Allison, T., A. Puce, and G. McCarthy. 2000. Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences* 4 (7): 267–78.
- Banaji, J. 2007. Personhood: Elusive but not illusory. *American Journal of Bioethics* 7 (1): 60–62.
- Bird, C. M., F. Castelli, O. Frith, and M. Husain. 2004. The impact of extensive medial frontal lobe damage on “Theory of Mind” and cognition. *Brain* 127 (4): 914–28.
- Churchland, P. S. 2007. The necessary-and-sufficient boondoggle. *American Journal of Bioethics* 7 (1): 54–55.
- H・トリラストラム・エンゲルハート 一九八八「医学における人格の概念」久保田顕二訳 『バイオエシックスの基礎』加藤尚武・飯田 亘文編 東海大学出版会 一九〇三二二
- Farah, M. J., and A. S. Heberlein. 2007a. Personhood and neuroscience: Naturalizing or nihilating? *American Journal of Bioethics* 7 (1): 37–48.
- . 2007b. Response to open peer commentaries on “Personhood and neuroscience: Naturalizing or nihilating?” Getting personal. *American Journal of Bioethics* 7 (1): W1–W4.
- Fendandez-Amesto, E. 2004. *So you think you're human? A brief history of humankind*. Oxford: Oxford University Press. (邦訳) フェリペ・フェルナンデス = アルメスト 『人間の境界はどこにあるのだろうか』長谷川眞理子訳 岩波書店 二〇〇八)
- Fratcher, J. 1979. *Humanhood: Essays in biomedical ethics*. New York: Prometheus Books.
- Freud, S. 2003. *The uncanny*. Trans. David McLintock. London: Penguin Books.
- 藤井直敏 二〇〇九 『ひまわりの脳』NOT出版
- Gallagher, H. L., and C. D. Frith. 2003. Functional imaging of “theory of mind.” *Trends in Cognitive Sciences* 7 (2): 77–83.
- Gong, L. 2008. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior* 24: 1494–1509.
- Grey, W., W. Hall, and A. Carter. 2007. Persons and personification. *American Journal of Bioethics* 7 (1): 57–58.
- 香川知晶 二〇〇九 『バイオエシックスにおけるゴキスター神話』『熊本大学生命倫理論集』生命という価値』高橋隆雄・糸和彦編 九州大学出版会 二一八―三三三
- 金森修 二〇〇七 『人とヒト——パーソン論の視座を通して』『シリーズヒトの科学』ヒトと人のあいだ』野家啓一編 岩波書店 一〇五―二二二
- Kanda, T., T. Hirano, D. Eaton, and H. Ishiguro. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction* 19 (1–2): 61–84.
- Kanwisher, N., and G. Yovel. 2006. The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society Biological Sciences* 361: 2109–28.

- Keysers, C., and V. Gazzola. 2007. Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences* 11 (5): 194–96.
- Kirsch, S., E. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3 (7): e2597.
- MacDorman, K. F., R. D. Green, C. C. Ho, and C. T. Koch. 2009. Too real for comfort? Uncanny response to computer generated faces. *Computers in Human Behavior* 25: 695–710.
- MacDorman, K. F., and H. Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7 (3): 297–337.
- Merker, B. 2007. Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30: 63–134.
- Meyers, C. 2007. Personhood: Empirical thing or rational concept? *American Journal of Bioethics* 7 (1): 63–65.
- 森政弘 一九七〇『不気味さの謎』『Energy』 七 三三三～三三五
- Morris, J. S., B. de Gelder, L. Weiskrantz, and R. J. Dolan. 2001. Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain* 124 (6): 1241–52.
- Moser, E., B. Derrn, S. Robinson, B. Fink, R. C. Gur, and K. Grammer. 2007. Amygdala activation at 3T in response to human and avian facial expressions of emotions. *Journal of Neuroscience Methods* 161: 126–33.
- 永守伸年 二〇〇七『脳科学と人格の概念』『実践哲学研究』 三〇 一四七～六一
- 信原幸弘 二〇〇九『心・意識・人への価値』『熊本大学生命倫理論集』 生命という価値』高橋隆雄・桑和彦編、九州大学出版会、八三～九九
- Owen, A. M., and M. R. Coleman. 2008. Functional neuroimaging of the vegetative state. *Nature Reviews Neuroscience* 9 (3): 235–43.
- Perrin, C. 2007. Against scientism, for personhood. *American Journal of Bioethics* 7 (1): 67–68.
- Peuskens, H., J. Vanrie, K. Verfaillie, and G. A. Ohsan. 2005. Specificity of regions processing biological motion. *European Journal of Neuroscience* 21 (10): 2864–75.
- Seyama, J., and R. S. Nageyama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16: 337–51.
- Steckenfinger, S. A., and A. A. Chazanfar. 2009. Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America* 106 (43): 18362–66.
- Sugrta, Y. 2008. Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America* 105 (1): 394–98.