

機械への倫理的行動

Ethical relationships between humans and machines

山本 吉伸[†]

Yoshinobu Yamamoto

[†]独立行政法人産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

yoshinov.yamamoto@aist.go.jp

Abstract

Is it possible to create a technology which leads us to ethical behavior for mass industrial products? When we asked the collaborators to give names to the machine, we have found that they were handling the named-machine with care. In this paper, we report on this experiment.

Keywords — Ethical relationships, Worldview, Uniqueness

1. はじめに

量産される工業製品たる機械が、人間にとって特別な存在、たとえば「応援する」「大切にする」「相手を守ってあげようとする」「相手のために自分の不利益をいとわない」といった、仲間同士、あるいはペットに対して日常的に観察できる倫理的行動の対象になれるかどうかは認知科学上の大きなチャレンジの一つである。人間が何に価値を見出すのかという議論をする上で重要な示唆を与えると期待されるだけでなく、産業上の観点からも興味深い。「機械を大切に扱う」という行動は大量消費・使い捨て文化の対極にあり、その行為を誘発する技術はサステナブルな社会の実現に貢献する。

「倫理的行動」を単純に定義することは困難であるが、本稿では「人間やペットのように置き換えできない存在（かけがえのない存在）として認識して接する」という意味で用いることにする。かけがえのない存在として認識する価値を、ここでは倫理的価値と呼ぶ。倫理的価値は対象の機能的価値に依存せず、対象の経済的価値や美術的価値その他の価値とも関係しない。

人間の倫理的行動によって保護される機械を作り

出す有望な方法は、人間と同様の機械を作り人間としてふるまってみせることであろう。人工知能研究はまさに人間の機能を持たせるアプローチであるし、ヒューマノイドは外観からのアプローチである[1][2]。

たしかに我々の行動は「対象が誰か」によって変容するのであるから、対象の外見が人間のそれであって機能的に人間と同じだとすれば、我々はそれらの対象に対して倫理的保護を与える可能性はある。しかし人間と区別のつかないアンドロイドが登場する未来を描いたいくつかの小説はそのようなアンドロイドたちが人間から迫害を受ける状況を描いており(例えば[3])、人間同様の外観・機能と倫理的な行動とが無関係であることは容易に想像できる。

2. 仮説

人間には、対象をかけがえのない存在であると認識する前提として「唯一性(uniqueness)」の認知があると推測される。工業生産品であっても「uniqueness」と認知させることは倫理的行動の前提となるはずである。加えて、個体を識別したとしても否定的な評価であれば倫理的行動を促進する必然性は見当たらないから、対象をなんらかの観点から肯定的に評価するきっかけとなる行為ないし事実が必要であろう。

それらを満たす方法として、我々は「命名」に着目した。

名前とは人間社会における最低限の参加資格の一つであり、名前のない構成員の存在は原則として認められていない。それに対して、多くの家電製品では製品番号と製造番号が付与されているものの、個々の個体に名前は付けられていない。

さらに、多くの文明・宗教で信任の情を表すために名前を授けるという風習がある。文学作品にも名前を与えるということが対人関係において重要な価値として取り入れられているものも少なくない。名付け親(Godfather)を名の親に次いで重要な対人関係として捉えることも多くの文化圏で見られる。名前を与えるという行為は、経済的価値や第三者からの評価、自己の技量向上をもたらすなど他の価値によって裏付けられるものではないにもかかわらず、多くの文明圏で肯定的な評価をもたらす行為と言える。

では、機械に対して名前を付けてもらったとしたら、その個体に対する態度に変化が起こるだろうか。

3. 実験

実験課題への協力に承諾した首都圏の大学 3 年生・4 年生 66 名に白紙を配布するとともに指定の Web ページへのアクセスを求めた。ランダムに 0 群 (34 名)、1 群 (32 名) にわけ、それぞれに別の設問を提示した。Web へのアクセスは任意のデバイスで実施してもらった。回答中は設問に集中してもらうために他者との会話を禁じた。Web 上では一問ずつ提示し、記述式回答を要求する設問には「最低回答時間」を設定し、一定の時間が経過しなければ次の設問に移動できないことを明示した。自由記述を求める設問では、問題に続いて「記述し終わったら「次の設問へ」のボタンを押してください。」「ここでは 10 分以上時間をかけて考えてください」「13 時 22 分まではボタンを押しても次の問題に進めません。」のように表示した。最低回答時間は設問によって 5 分の場合もあったがその他は同じように提示された。

共通提示 (プロローグ)

以下のプロローグは写真(図 3-1)とともに全実験協力者に共通に提示された。

あなたは T 市にある機械メーカーの工場に見学に来ました。伯父さん (父親の兄) が工場長をしていたので、見学させてもらうことになったのです。

この工場では 2000 台の特別仕様の FAX を製造していました。

この 2000 台はそれぞれ顧客に配送されましたが、コンピュータのミスで 2001 台が製造されていたため、引き取り手のなかった一台だけが倉庫に残っていました。



図 3-1 FAX

工場見学に来ていたあなたは、工場長である伯父さんからこの一台を貰うことになりました。

FAX の機能

外観は普通のオフィス仕様ですが、この特別仕様の FAX には、「あなた宛ての FAX が届きました」など、音声でしゃべる機能が付いています。また、利用者の音声を聞き取る機能が付いています。FAX 回線を通じてサーバから新しいソフトウェアを自動的にダウンロードしてきてアップグレードしたり、利用者の音声や操作から学習したりする機能が付いています。

さらに、学習機能によって独自の機能を持つようになっていきます。話しかけたり、長く使ったりしているとだんだん他の FAX とは違ってくる仕組みになっています。

FAX の外観

届いた FAX は、ほとんど気づかないぐらいのものであったが受話器の位置に薄い赤いシミのような色がついていました。

0 群への教示 (非命名群)

プロローグに続き、0 群には以下の設問が提示された。

設問 1 未来の FAX

この FAX にもないような、未来の FAX に装備さ

れている機能を考えてください。

その機能がどんなものか、その機能がどんな時に役立つのか（あるいは役立たないエピソード）などを自由に創作してください。利用環境など自由に設定していただいて結構です。登場人物はすべて架空の人物でも実在の人物でも構いません。

設問2 月日が流れ・・・

使い始めてから5年が経過したとします。このFAXの学習機能により、どのような性能向上がなされているかを想像して書いてください。また、他と違ってきているところがどこかなど、「5年間経過後のFAX」について記述してください。

設問3 自分の名前

あなたの名前の由来を教えてください。

1群への教示（命名群）

1群には、プロローグに続いて以下の設問が提示された。

設問1 名前を付けてあげてください

まずこのFAXに、名前をつけてあげてください。ほかのFAXと同じにならない、特別な名前を考えてあげてください。

その名前を付けた理由や、その名前を付けるに至ったエピソードを創作してください。FAXをもらったのがいつごろだったのかも自由に設定していただいて結構です。登場人物はすべて架空の人物でも実在の人物でも構いません。

設問2 月日が流れ・・・

それから5年が経過しました。***はどんなところが最初のときから変わりましたか？また、どんなところが他と違ってきているところですか？***をくれた伯父さんに「その後の***」を報告する手紙（文章）を書いてください。

¹ 設問1で設定した固有名詞が***に埋め込まれて表示された。

設問3 ***の学習

***は、印刷の用紙がなくなったとき最初「用紙がなくなりました」と発声していたのに、いまでは「おなかすいた！おなかすいた！」と発声しています。なぜ変わったのか、理由や経緯を教えてください。

条件01 共通提示（主質問）

以下、両方の条件群に図3-2とともに提示した。

設問4 その後

もらったFAXの今の写真です。なにがあったのか、理由を教えてください。



図 3-2 壊れた FAX

設問5 修理

工場に問い合わせたところ、FAXの修理は可能ですが、学習データは失われる可能性もあるそうです。

限定生産のため、同じ機種はもう入手できないそうです。希望すれば、量販店で売っている一般的なFAXに無料で取り替えてくれるとのこと。

あなたなら、どうしますか。選択してから「次の設問へ」のボタンを押してください。

(0) 修理してほしい

(1) 新品に取り替えてほしい

以上に加え、「ここまでの記述式だった各回答で、それぞれ何文字記述したか」を回答してもらった。

4. 結果

図4-1は0群（非命名条件）と1群（命名条件）で、設問5にどのように回答したかを比較したものである。

0群と1群には有意差が認められた($p<.005$). すなわち**機械に命名した人はその機械の修理を望み、機械に命名しなかった人は機械の交換を望んだ。**

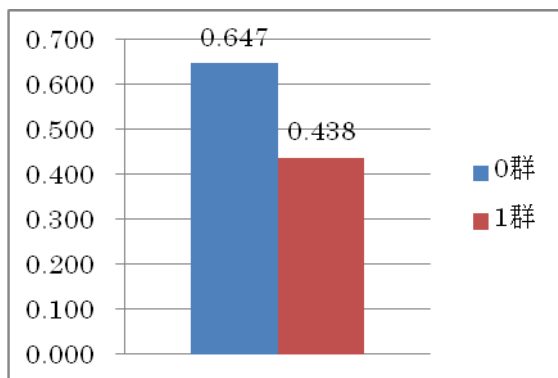


図 4-1 設問 5 の平均

5. 考察

FAX は、外観はまったく人間とは異なるものであり、なおかつ名前を付けている人は少ないと考えられる工業量産品(同一外観の機械が量産されている)の典型例である。写真で示される FAX には特徴的な外観はない。

それにも関わらず、簡単な教示条件(名前を付けてもらう)だけで機械の交換を避けて修理による対応を望む反応(倫理的行為)が引き出された。

5.1 効果の源泉

自分で名前を付けたものに対しては愛着を感じる。この経験則は、具体的に応用されて成果をあげている。サービス計測システムの導入実験[4][5]を実施している我々の研究センターでは、実験プロジェクト終了後も継続してサービス計測システムを利用してもらえるようシステムへの愛着を高める手法として「システムに名前を付けてもらう」ことを実施している。城崎温泉ではプロジェクト終了後もシステムが継続利用されている(ただし、命名の効果だけで継続利用されているわけではないと思われる)。

車や人形などに名前を付ける人は散見される。これは「愛着を持っているから名前を付けるのではないか?」とも考えられる。この点、本実験は、もともと愛着を持っていない機械に対して名前を付けて

もらう状況を扱ったものであるから、因果関係に関しては「命名行為によって愛着を感じた」と指摘しうる。この結果は「愛着があるから命名する」という仮説を否定するものではない。

本実験の結果から、命名には具体的な効果が認められるものの、命名行為は(対象物に倫理的価値を認知する上で、より本質的な「何か」の)誘引になっているだけではないか、と考えることもできる。名前がなくとも、愛着や倫理的価値を感じる事例は無数に存在しうるからである²。

また、能動的に命名することの意味も問題となる。人間やペットなどをかけがえのない存在と認知して愛着を寄せるとき、多くの場合は自分で命名してはいない。近年では、「はやぶさ」の例³が記憶に新しい。「はやぶさ」は JAXA による命名であるのだから、自分で命名することと愛着を持つこととは関係ないようにも思われる。

ただ、注意すべきは、はやぶさは一つの特別な個体として認知されていた点であり、その点で工業量産品と相違する。特別な愛着を感じる以上、少なくとも他の個体と識別して認知されることが前提だとすれば、命名は個体を識別するという点に効果がある可能性がある。もしこの仮説が正しければ、他から与えられた名前を使う(またはあらかじめ用意された名前群から一つを選ぶ)方法と、自分で命名する方法とで差がでないことになろう。能動的に名前を付けることにどのような意味があるのかの議論は今後の課題である。

なお、命名時に実験協力者に「その名前を付けるに至ったエピソードを創作してください。」と教示しているが、これはあまり考えることなしに命名されてしまうことを避ける趣旨で設けられている。単に

² たとえば Ralph Steadman 著「The Little Red Computer(邦題:らくだいコンピューター;偕成社)」では、主人公だけ赤く塗られて他と区別される他、名前はない。

³ 2003年5月9日に打ち上げられた小惑星探査機。60億 km の旅を終え 2010年6月13日地球に大気圏再突入して本体は消滅した。その際、日本では探査機を擬人化して盛り上がりを見せ、幅広い世代が探査機の帰還を応援した[6]。

名前をつけてくださいという課題だと「FAX 君」のような名前が多数を占めることになる。このような名前では個体識別の趣旨は没却される。

5.2 世界観の創作

では仮に、命名行為そのものは個体識別の意味だったとすると、個体に対して倫理的な価値を付与させた可能性としてなにが考えられるか。

ここで、「世界観(WorldView)」モデルを考える。

世界観とは世界(外界)についての統一的な見方・考え方のことである。外見や機能がいかに人間と同様であったとしても、対象に倫理的価値を見出すかどうかは「対象を自分の仲間としてみなすかどうか」にかかっている[7]。対象が「敵」や「奴隷」であるコンテキストでは、外見や機能はまったく同じ人間にもかかわらず(実際、人間である)、我々は容易に非倫理的行動をとることができる。逆に、対象が「人間の仲間」というコンテキストでは、イルカやシャチに対しても倫理的行動をとることになる。戦場カメラマンにとって、苦勞を共にしたカメラは「戦友」すなわち仲間として倫理的行動の対象となることもあるし、自分の子供に対しても非倫理的行動が発現する世界観もありうる。しかも人間は単一の世界観だけで活動するのではなく、適宜世界観を入れ替えることができている。世界観は体験の共有によっても形成されるだろうが、体験に限られない。

本実験では、実験協力者自身が世界観を構築した

と考えられる。それは、設問3にある「物語の創作」である。本実験の0群と1群とでは、命名行為の指示以外に異なる点でもある。

ここでの物語とは、当該個体にだけ成立するエピソードのことである。

たとえば、物語をより多く創作したほうがより多くの影響を受けるのではないか。逆に、創作課題にあまり熱心とは言えない人(創作量の少ない人)ほど、物語の影響を受けないのではないか。同様に、課題に取り組む時間が短いほど物語の影響を受けず、課題に取り組む時間が長いほど物語の影響を受けるのではないか。

命名群(1群)を詳細にみてこの点を確認する。

図5-1は文章量と課題の回答時間に着目したグラフである。命名群のうち、回答文章量が平均より多い人を1HB群、平均より短い人を1LB群とする。命名群のうち、回答文章量が上位5人を1H5B群、下位5人を1L5B群とする。命名群のうち、回答総時間が平均より多い人を1HT群、平均より短い人を1LT群とする。命名群のうち、時間が上位5人を1H5T群、下位5人を1L5T群とする。グラフは図中の凡例順に左側から並んでいる。

回答文章量が多ければ、それだけ物語について創作量が多いと考えられるので、1H5B群はより修理希望者が多くなると予想される。しかし今回の実験

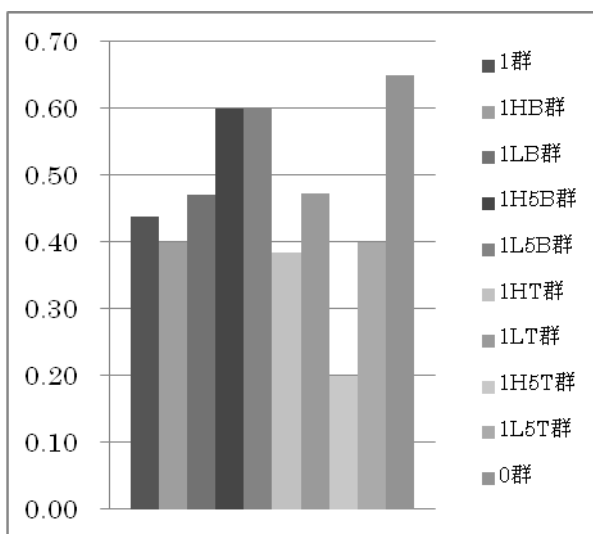


図 5-1 文章量と課題の回答時間

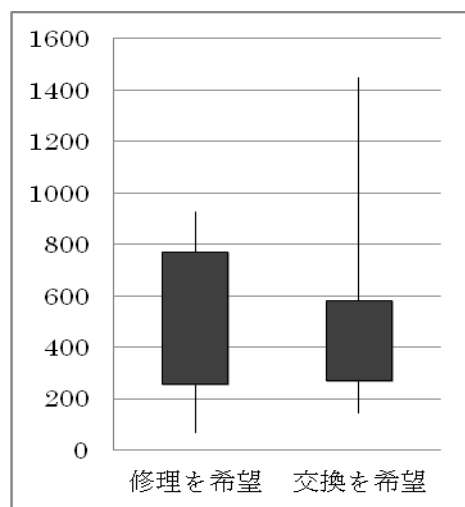


図 5-2 条件 1 群の回答文章量

ではむしろ新品への交換希望者比率が高くなっている。一方、回答文章量が少ない 1L5B 群は修理希望者が少なく、こちらは仮説を支持するようにみえる。

修理を希望した人と交換を希望した人のそれぞれの回答文章量を比較したものが図 5-2 である。相対的に文章が長い。

一方、回答時間が短ければ（早く終了すれば）それだけ物語を検討する時間も少ないと考えられるので、1L5T 群は修理希望者が少なくなると予想される。ところが今回の実験では修理希望者が多くなっている。一方、回答に時間をかけた 1H5T 群は高い割合で修理を希望しており、こちらは仮説を支持するようにみえる。

1H5T 群も 1L5T 群も修理希望者が多い本実験の結果は仮説を支持しない。設問への回答時間は、設問の提示から次の設問に移るまでの時間であるから、必ずしも設問について検討していた時間の長さではない。同様に、1H5B 群と 1L5B 群はいずれも新品への交換を希望する比率が高かったのであるから、回答時間・回答文章量の長短で傾向をみることはできなかった。

物語仮説が機械に対する認識に影響を与えていると解釈するのであれば、それは物語の質的問題と密接に関連せざるを得ず、客観的指標だけでこれを捉えることには限界があるようにも思われる。また、文章量は課題依存より個人の力量や性格にも大きく依存するので、それらの分析なしに客観的指標として世界観との関連を論じることは困難と想像される。

なお、プロログ中で FAX の外観に触れているが、これは創作しやすさの一助となることを期待して付加したものであったが、これに言及した回答は 2 件だけであった。

6. まとめ

本稿では、機械に対して倫理的行動をもたらす技術を模索する目的で、機械に名前を付ける実験を実施したところ、名前をつけた実験協力者群で多くの倫理的行動が見られた。

単に機械を「大切に扱う」ことが目的であれば、たとえば経済的価値に着目した観点から議論する（壊すと経済的サンクションを受ける）ことも可能であるが、本稿では経済的価値では説明し得ない利用者の行動について検討した。

工業量産品たる機械は、同機能のものが再生産可能であり、そのことを認識している成人実験協力者の性格によらず、効果のある技術として、名前をつけてもらう方法は有望であることがわかった。しかし、名前を能動的に付けることがもつ意味などは今後の課題として残されているし、この実験が現実世界の現象や擬人化[8]などの事例をどの程度説明できるかも今後の課題である。我々は世界観モデルを導入してこれらの検証を継続する予定である。

ただ、少なくとも本実験の結果が強く示唆するのは、人間以外の存在（たとえばクジラ）に対して保護の意欲を高めようとするならば、暴力的手段によって訴えるより、クジラに名前をつけてこれまでの半生の記録を報道するほうが効果的であろう、ということである。

参考文献

- [1] 土肥浩, 石塚満, (2002) "WWW と連携する擬人化エージェントとの HAI", 人工知能学会誌 vol.17, No.6, pp.693-700.
- [2] Byron Reeves, Clifford Nass, (1998) "The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places", Cambridge University Press,
- [3] Philip K. Dick, (1968), "Do Androids Dream of Electric Sheep?", Del Rey.
- [4] 山本吉伸, 北島宗雄, (2011) "オープンサービスフィールド型 POS による観光客動向把握の技術", 観光情報学会誌, No.7, No.1, pp.47-60.
- [5] 中島正人, 福原知宏, 三輪洋靖, 西村拓一, (2012) "介護サービスにおける申し送り支援システムの開発" モバイル学会誌, (印刷中)
- [6] <http://www.youtube.com/watch?v=jSpZHW0DvNk&feature=related>
- [7] 山本吉伸, (1993) "ロボットとの対話", 人工知能学会研究会, SIG-J-9302-3, pp.327-334.
- [8] Nicholas Epley, Adam Waytz, and John T. Cacioppo, (2007) "On Seeing Human: A Three-Factor Theory of Anthropomorphism", Psychological Review, Vol.114, No. 4, 864-886.