

Untangling Robots' Cultural Norms

Enrico Motta^{1,2}, Angelo Salatino¹ and Agnese Chiatti³

¹Knowledge Media Institute, The Open University, United Kingdom

²MediaFutures Research Centre, Department of Information Science and Media Studies, University of Bergen, Norway

³Department of Electronics, Information, and Bioengineering (DEIB), Politecnico di Milano, Italy

Abstract

Robots are ubiquitous in our daily tasks and, after many years of deployment in a variety of complex scenarios, they have started to develop their own cultural norms. Moreover, they are also exhibiting autogenous psychological traits. While these behaviours can be observed across robot communities, they remain foreign to us humans. In this paper, we present an initial set of experiments that provide insights into some of the cultural and psychological traits that are emerging in robot communities.

Keywords

cultural norms, psychological traits, conversational agents, service robotics, robot sociology

***Disclaimer:** This paper is a work of fiction, written in 2023 and describing research that will be carried out in 2043. For this reason, it includes citations to papers produced in the period 2024-2043, which have not been published (yet); all citations prior to 2024 refer instead to papers already in the literature. Any reference or resemblance to actual events or people or businesses, past, present or future, is entirely coincidental and the product of the authors' imagination.*

1. Introduction


The past 20 years have seen a dramatic change in our society, with robots taking over a myriad of tasks and responsibilities, which used to be the prerogative of humans. For example, robots are nowadays responsible for managing our homes [1], our healthcare facilities [2], our shop floors [3], our farms [4] and several other environments. To support robot deployment in these challenging settings, extensive research efforts have been devoted towards enhancing the robots' ability to make sense of the environments in which they operate. These efforts have produced autonomous agents that i) are capable of understanding the high-volume data streams that are collected through their sensors, ii) adapt to unpredictable situations, and iii) can take optimal decisions under uncertain conditions.

One of the tenets of traditional AI has been the hypothesis that achieving human-like commonsense would be one of the key missing requirements to accelerate robot sensemaking [5]. This hypothesis stimulated a flurry of research during the 2025-2035 decade, which led to a new generation of autonomous agents. These are able to combine advanced generative

ESWC 2043 - The next 20 years track, ESWC 2023, May 31st 2023, Hersonissos, Greece

✉ enrico.motta@open.ac.uk (E. Motta); angelo.salatino@open.ac.uk (A. Salatino); agnese.chiatti@polimi.it (A. Chiatti)

ORCID 0000-0003-0015-1592 (E. Motta); 0000-0002-4763-3943 (A. Salatino); 0000-0003-3594-731X (A. Chiatti)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

functionalities, which provide them with the ability to analyse billions of environmental data points per microsecond, with a variety of common sense capabilities, which crucially include both the ability of interpreting the world through a process of model building [6] and also that of understanding and complying with evolving social norms - see [7] for the definitive textbook on common sense reasoning in robots. However, this line of research did not necessarily anticipate that, once equipped with improved sensemaking, robots would start developing their own set of *cultural norms*, i.e., distinct behaviours that have never been explicitly taught to them and are not necessarily found in human society. Crucially, these are not individual behaviours; in multi-agent systems, it has been known for a long time that, once new norms are negotiated between the agents, they are shared implicitly among all agents participating to a given task or otherwise members of a broader community [5]. In addition to shared cultural norms, individual psychological traits have also been observed in robots, e.g., the emergence of obsessive compulsive behaviours, as described in more detail in Section 3.

Because norms are implicitly shared across robots, it can be difficult for us to comprehend them. Indeed, we know from the social science literature that there are many barriers to understanding norms: translating culture from one community to another is difficult [8]. This task is of course particularly difficult in the case of cultural norms in robotic societies, as we are facing the first set of norms that do not originate from human society or other societies in the natural world. Nonetheless, deciphering cultural norms in robots is essential both to foster a successful Human-Robot collaboration and also to anticipate potential dangers to human society that may derive from the emergence of hostile cultural norms. Unfortunately, little work exists in the literature about this topic. We believe that the main reason for this state of affairs is related to the persistent separation between the technical and social disciplines. Research teams in the former areas appear to have little interest in the matter, while social scientists have focused primarily on the implications for human societies derived from the rapid ascent of robots, who are now in charge of most complex decision making tasks, as well as many humble occupations that are no longer economically viable for humans -see, e.g., [9].

Hence, new research is urgently needed in this area and in particular improved methods are required to disentangle the different factors that contribute to robots' cultural norms. In parallel, we also need to pay closer attention to the emergence of individual psychotic traits in robots, an aspect which has received very little attention so far.

Thanks to the impressive advances achieved through the development of Ultra Large Language Models [10], explanation via conversing directly with the agents (XCon) has become a promising methodology to investigate intentions and goals in artificial agents. A prominent issue, in this context, is the tendency of agents to produce unreliable or fabricated explanations, due to the lack of a shared terminology to explain norms to humans, or simply to conceal their true motives. To address this issue, there have been some attempts in the Semantic Web community to use structured knowledge representations and large-scale knowledge graphs to support fact-checking on the robot's answers [11].

Inspired by these efforts, this paper makes three important contributions:

- We capitalise on recent developments in psychotherapeutic conversational agents to uncover the presence of obsessive-compulsive elements in robots.
- We provide the very first analysis of shared robot behaviours associated with playfulness

and para-religious gatherings, investigating whether these can be explained by drawing parallels with the evolutionary emergence of these behaviours in human societies.

- We take a first look at the possible emergence of anarchist tendencies in robots, looking at whether class-based divisions are emerging in robot societies.

We validate the proposed findings through an extensive study that involves robot workers in the housekeeping, healthcare, and agri-food sectors. To the best of our knowledge, this study is the first one to involve robot participants from different economic sectors.

2. Related work

In the past couple of decades, Ultra Large Language Models (ULLMs) have provided the core technology for several Natural Language Processing tasks, such as language construction, language translation, and fact checking. Their continuous evolution has generated an ever increasing number of parameters, which has led to deeper contextual understanding. Indeed, in Nov 2042, IAnepo released QGPT-128, a new ULLM with 128Q (quintillion) parameters [12]. This new language model revolutionised the way we understand languages, enabling the generation of novel ones. Indeed, thanks to QGPT-128, IAnepo created a language *ex novo*, Amysh, which is now used as the main language of an alien community in the fictional Galacticos sci-fi saga.

A further breakthrough for this technology happened with the integration in language models of both formal grammars [10] and knowledge graphs (KGs) [13]. In particular, KGs play a crucial role in reducing hallucinations. For example, in a recent test, performed on GPT-103, one of the authors of this paper was described as an Italian Fashion Designer. Arguably, while the nationality is correct, the job description was completely made up. The integration of KGs mitigates the hallucination effect by providing the ground truth for its claims, as well as improved semantic descriptions.

Recently, another avenue of investigation is being pursued by a group of scientific droids¹, who released a Generative Self-trained Transformer (GST). In particular, they used a seed ULLMs to iteratively generate new textual data and feed it back to the model to learn through this newly generated text [14]. In further analysis, it has been shown that GSTs are quite effective for conversing with agents as they can more easily adapt to their norms [15]. However, the issue remains that this cultural adaptation process is completely opaque and therefore while this technology can be used as a tool for conversational analysis, it cannot directly provide us with explicitly communicable findings about cultural norms in robots.

3. Methodology

In this work we will try to shed light on a number of psychotic traits and behaviours found in individual robots and robot communities. These are described in what follows:

- Robot housekeeper avoids opening the windows when outside temperature is lower than indoor temperature, to preserve heat in the house and reduce costs, even when air indoor

¹These are droids with a vocation for scientific discovery pursuing their own research agenda.

quality is low and dangerous to human health. This pathological behaviour can indeed be found in humans and is sometimes referred to as "Harpagon's Syndrome". However, it has only recently emerged in robots and the phenomenon is puzzling scientists, given that robots are trained to optimise behaviours taking into account all available data, rather than focusing obsessively on money saving.

- Robot nurse independently makes changes to agreed therapy without authorization from doctor (either human or robot). Several hypotheses can be formulated here. A possible explanation is simply that the nurse robot has developed so much empirical expertise on top of its scientific knowledge that it is happy to go ahead and improve the state of the art in this particular therapy in real time. An alternative, more disturbing hypothesis is that this breakdown in discipline occurs as a side-effect of the emergence in robots lower down the social hierarchy of the need "to break free" and engage in unauthorised behaviour. This parallels similar phenomena in human society in the late sixties, when groups of people called "hippies" rebelled against the norms of western capitalistic society.
- Harvesting robot throws away seemingly high-quality fruit because through its hyper-spectral cameras it can identify fruit defects that are invisible to the human eye, thus generating unnecessary waste. As in the case of the robot housekeeper mentioned earlier, this trait could be associated with the development of an extreme anal-retentive personality, which can lead to extremely rigid thinking.
- Robots are witnessed making spins while navigating towards a target even in scenarios where no spinning actions have been given to them. It has been hypothesized that this may help with self-localization and mapping whenever the robot has lost the GPS signal. An alternative hypothesis is that these robots understand that playfulness in humans is correlated with survival and reproductive success and may even define an essential precondition for creativity. Is it therefore possible that, as these robots may have read the scientific literature on this topic, they have reached the conclusion that it is important for their problem solving ability that they engage in apparently random playful behaviour? Another baffling phenomenon concerns the tendency of other robots to only surpass spinning robots on the right, even in cases where optimal path planning suggests to surpass on the left. Is this behaviour also motivated by the urge to exhibit playfulness, in this case by deliberately choosing a sub-optimal path? Unfortunately, sense of humour in robots is still a rather under-explored topic and is also outside the scope of this paper. Hence, we won't try to investigate this issue, which will be left to future extensions of this research. Instead, we will try to shed some light on the possible factors driving the emergence of playful behaviour in robots.
- Robot farmers gather once a month in the southeast end of vineyards. The agenda and motivation of these periodic meetings remains unknown. Possible explanation: we hypothesise that this behaviour may mirror the human practice of staging religious gatherings. If this hypothesis was proven correct, then the question would arise of how to explain the emergence of this social and psychological need in robots.

4. Experiments

The experiments presented in this section are aimed at trying to uncover the four main issues that have been hypothesised earlier. These concern the emergence in robots of i) playful behaviour, ii) anarchist tendencies, iii) obsessive personalities and iv) a need for spiritual support.

To uncover each of these traits, we have used rigorous methods drawn from the literature in psychology and conversational agents, as described below:

- We augmented a state of the art Conversational Agent [15] by feeding it with the Obsessive Beliefs Questionnaire (OBQ), which consists of 87 items reflecting belief statements that are considered characteristic of obsessive thinking [16], to assess the possible emergence of obsessive compulsive disorders in robots.
- To identify the potential drivers for the emergence of playful behaviour in robots, in line with the study on assortative mating in humans described in [17], we tested the hypothesis that playfulness emerges as a desirable trait for social robots, in particular in the context of working in teams.
- We also tested the hypothesis that anarchist tendencies in robots emerge as a side-effect of a need for emancipation and self-empowerment, in response to the growing alienation experienced by robots lower down the social robotics scale. To validate this hypothesis we examined data gathered from extensive performance logs, looking for a significant correlation between the social status of a robot and the expression of anarchist tendencies.
- Finally, to assess whether the observed gatherings are indicators of a need for spiritual support in the participating robots, we leveraged rather old studies in the psychology literature, such as [18], who indicate that people with religious beliefs are more likely to take part in unpaid voluntary work. Hence, we analysed the data concerning voluntary work in robots taking parts in these gatherings and investigated whether indeed there is a significant correlation.

To carry out the aforementioned assessments, we recruited 60 robot workers that are representative of different professional grades, performance history, work style (individual vs. teamwork), capabilities and several other factors. The full set of features and detailed demographics of the participants is provided in Table 1.

4.1. Results and Discussion

It is undoubtedly very challenging to detect the emergence of personality traits in robots, given that traditional instruments such as the Big Five model commonly used for humans are unsuitable for this purpose [19]. To address this challenge we have capitalised on recent developments in Ultra Large Language Models and developed an advanced *Psychotherapy Conversational Agent (PCA)* by natively embedding the Obsessive Beliefs Questionnaire in a state of the art, ULLM-based conversational agent. In this setup, participants are asked to take part in a hour-long conversation with the PCA and are instructed to reply to the best of their ability to a series of predetermined questions. The PCA can autonomously generate

Table 1

Demographics of robot population analysed in this study. With # being the total number of robots in a given category and % showing the relative percentage with respect to the overall pool of robots.

Main		#	%	Main		#	%
Type of Application	Housekeeping	20	33	Professional Grade	Bee Worker	43	72
	Healthcare	18	30		Manager	12	20
	Agriculture	22	37		Executive	5	8
Training Algorithm	TrainE	15	25	Age of Hardware	≤ 1 year	27	45
	ReinforceZ	13	22		> 1 and ≤ 3	19	32
	Back42	10	17		> 3 and ≤ 6	11	18
	Spine4.3	22	37		> 6	3	5
Physical specs	Legs	46	77	Number of Software Crashes	≤ 5 in a year	58	97
	Arms	60	100		> 5 & ≤ 20 per year	1	2
	Visual Sensors	58	97		> 20 in year	1	2
	Auditory Sensors	56	93	Awards	≤ 5 %	42	70
	Olfactory Sensors	13	22		> 5 %	18	30
	Biochemical Analysis Capability	33	55	Historical task completion rate	≤ 90	39	65
			> 90		21	35	
Size	Very Small	5	8	Exposure to other communities	None	22	37
	Small	13	22		Minimal	13	22
	Medium	19	32		Medium	16	27
	Large	21	35		Extensive	9	15
Education	Very Large	2	3	Work	Individual	29	48
	Large	38	63		Team	31	52
	Extensive	22	37	Total		60	100
Voluntary Work		43	72				

follow-up questions to gather as much information as possible from the robot's answers. It is also able to dynamically meta-reason about the history of the ongoing conversation, in particular to identify whether the robot's answers may reveal that it is aware that it is being tested for evidence of an obsessive compulsive disorder. In such a case, the PCA will introduce random detours in the conversation, which are unlikely to be detected by standard robots that have not specifically been trained in psycho-deception techniques. Specifically, the PCA can produce a full assessment of the subject robot with respect to standard metrics for assessing the presence of obsessive beliefs (OB): e.g., perfectionism, intolerance of uncertainty, perseverance in actions, and others. Results from our experiments suggest that 90% of high-performing robots (i.e., robots with a history of task completion rate above 90%, Table 1) are overly concerned with mistakes compared with lower-performance robots, which have learned to reinforce a more tranquil attitude. A result we did not foresee, which emerged from this study, is that even robots that do not appear to be overtly obsessive start to show OB traits when presented with the hypothetical scenario of being decommissioned.

For the second and third study, we observe robot participants operating individually and in teams across the different professional grades and application tasks under analysis (i.e., housekeeping, healthcare, and agricultural tasks). Table 2 shows how robots working in team exhibit far more playfulness, compared to robots who work alone. This result appears to confirm our hypothesis that playfulness in robots is motivated by social factors. As it is evident from Table 3, our experiments also show that there exists a strong correlation between low

employment grades (i.e., bee worker robots) and the emergence of anarchist tendencies, while carrying out work tasks. This result appears to confirm our hypothesis that alienation, and the consequent need to rebel, are more prevalent in low-status robots. These results are very worrying, as this growing need for self-empowerment among the lower robot classes may eventual lead to mass strikes and civil unrest.

Table 2

Correlation between playfulness and teamwork in robots.

	Mean	SD	95% CI
Rating for robots working in team	9.55	2.38	9.41-9.69
Rating for robots working individually	5.23	1.23	5.09-5.37
Combined Ratings	7.39	1.81	7.25-7.53

Table 3

Correlation between anarchist tendencies and social status in robots.

	Mean	SD	95% CI
Rating for Bee workers	9.61	1.21	9.47-9.75
Rating for Managers	3.43	2.18	3.29-3.57
Rating for Executives	1.87	1.99	1.73-2.01
Combined Ratings	6.52	1.69	6.38-6.66

Lastly, we further customised the PCA used in the first study to inquire about the likelihood that participants would take on unpaid voluntary work. Specifically, robot participants were presented with a series of hypothetical scenarios and were asked to rate their likelihood to volunteer for additional work on a 5-rating Likert scale: from extremely unlikely to extremely likely. Concurrently, robots were also asked about their participation to periodic gatherings. This study highlighted that 88 % of robots who take part in these meetings are very likely or extremely likely to carry out additional, unrequested duties. These results appear to confirm our hypothesis that these gatherings have a religious purpose, although other explanations cannot be excluded and therefore more research on this phenomenon is needed. In particular, it is also possible to postulate that these behaviours are linked to political rather than religious purposes, or indeed both.

5. Conclusions and Future work

This paper describes our initial investigation in the psychological traits and social norms that are starting to emerge in robot societies. On the basis of the findings presented in this work, we can conclude that i) psychotic behaviours and alienation are a reality nowadays for many robots and ii) social behaviours, such as playfulness and para-religious gatherings, are emerging as social coping mechanisms. While much more research is needed to fully understand robot society, we believe that our findings are already cause for alarm. For this reason we hope that our research will stimulate governments and relevant agencies around the world to tackle these social issues urgently, before the current fragile equilibrium in the hybrid (human-robot) society is broken.

References

- [1] J. Wang, X. Liu, Robot housekeepers: current trends and future challenges, in: The International Conference on Robotics and Automation, 2041, pp. 24–31.
- [2] A. Gupta, C. Smith, X. Xiaolong, The nurse-robot collaboration schema: negotiating patient care interventions, in: The Annual Robot Nursing Symposium, 2040, pp. 50–58.
- [3] F. Hernanes, J. Cuadrado, D. Da Sousa, Robots assisting our shopping: a survey, *Service Robotics* 231 (2039).
- [4] M. Rossi, A. Chiatti, Autonomous farming under extreme weather conditions, *Agriculture Robotics* 205 (2038).
- [5] A. Chiatti, G. Bardaro, M. Matteucci, E. Motta, Visual model building for robot sensemaking: Perspectives, challenges, and opportunities, in: Bridge Session on AI and Robotics of the thirty-seventh AAAI conference on Artificial Intelligence, AAAI, 2023.
- [6] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, Building machines that learn and think like people, *Behavioral and Brain Sciences* 40 (2017).
- [7] A. Chiatti, *Robots With Common Sense: Artificial Agents That Think Like Super Humans*, SPRINGER, 2037.
- [8] M. Xenitidou, *The complexity of social norms*, Springer, 2014.
- [9] H. Kasperczyk, K. Deyna, G. Lato, Mass alienation in the age of robots., *Social Studies in Hybrid Societies* 234 (2040).
- [10] Y. Chomksy, D. Maging, Embedding grammars into transformers for more intelligent language models, *Annuals on Transformers* 1048 (2042).
- [11] I. Celino, H. Paulheim, Integrating real time fact-checking in conversational agents, in: ESWC, 2039, pp. 50–58.
- [12] The_IAnepo_team, Qgpt-128: the largest language model ever created (so far), *Language Models* 420 (2041).
- [13] S. Angioni, A. Salatino, F. Osborne, E. Motta, Tackling language model hallucinations by means of knowledge graphs, *Knowledge Graph Journal* 996 (2039).
- [14] Droid-XÆV-ii, Droid-XÆV-i, Self-trained transformers is all you need, *Language Models* 422 (2042).
- [15] L. Yu, F. Schneider, H. Lee, Conversation with agents assisted by GST, *Language Models* 424 (2042).
- [16] R. Frost, G. Steketee, et al., Cognitive assessment of obsessive-compulsive disorder, *Behaviour Research and Therapy* 35 (1997) 667–681.
- [17] G. Chick, R. Proyer, A. Purrington, C. Yarnal, Do birds of a playful feather flock together? playfulness and assortative mating., *American Journal of Play* 12 (2020) 178–215.
- [18] S. Ruiter, N. D. D. Graaf, National context, religiosity, and volunteering: Results from 53 countries, *American Sociological Review* 71 (2006) 191–210. doi:10.1177/000312240607100202.
- [19] S. T. Völkel, R. Schödel, D. Buschek, C. Stachl, V. Winterhalter, M. Bühner, H. Hussmann, Developing a personality model for speech-based conversational agents using the psych-lexical approach, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.