

Social Robotics and the person problem

Stephen J. Cowley¹

Abstract. Like computers before them, social robots can be used as a fundamental research tool. Indeed, they can help us to turn our attention from putative inner modules to thinking about the flow and emergence of human intellectual powers. In so doing, much can be gained from seeking solutions to MacDorman’s *person problem*: how can human bodies – and perhaps robot bodies – attune to cultural norms and, by so doing, construct themselves into persons? This paper explores the hypothesis that social robots can be used to ask fundamental questions about the nature of human agency.

For social robots to live up to their name, the focus needs to fall on functional co-ordination and co-action. This enables one to link research on how today’s robots function as social mediators with engineering approaches that explore both how understanding can be hard-wired, how this influences the cultural ecology and, perhaps, in designing robots that can discover how we enact values. To do this new kinds of collaboration need to be established. The key theoretical question is whether, in becoming persons, humans depend on embodiment alone or, as suggested here, intrinsic motive formation enables them to discover the distributed forms of embodiment favoured by culture.

1 INTRODUCTION

Since social robotics is in its infancy, no-one knows what impact these machines will have. Especially in the West, their potential is apparent to few. I felt it – and I mean felt it – when I went to Osaka to talk about infant development. There I was introduced to androids; my hosts suggested that, in realizing their power, I might help by proposing an engineer friendly model of human behaviour. What follows is, in part, a polite refusal. Rather than build robots that simulate what psychologists *say* humans do, social robots can free us of mentalist fantasies. To do this, I argue, they must be recognised as a research tool. Their value arises in investigating the real-time flow and changing results of joint action. Eschewing appeal to psychological competencies, engineering design can thus co-emerge with observational studies of humans and systematic investigation of how a cultural ecology adjusts to social robots.

Co-action can be defined as occurring when “one agent’s action is influenced by or occurs in the context of another agent’s– and together they do something that is not fully attributable to either one alone” [50]. As explained below, robots with this capacity can *simulate* understanding of what happens between people. Given that they already produce social affordances, this opens many new applications. This paper however, focuses on other questions. Just as computers showed us about cognition, it is argued, robots can be used to understand

human forms of minded behaviour. As MacDorman and Ishiguro emphasise, robots can be a test-bed for fundamental research [33]. Whereas computers clarified thinking about *mind*, robots enable us to formulate and test hypotheses about human agency. By contributing to social encounters, they can throw light on how physical and cultural resources impact on what we think, feel and do. By asking bold questions, robotics may lead to syntheses that link the work of engineers, psychologists, philosophers and social scientists.²

2 BEYOND COMPETENCIES

Since Chomsky’s work in the 1960s, the engineer’s universals have dominated models of human powers. Notoriously, even language is often associated with an innate module. As a result of a *separatist* approach, cognitive science has faced difficulties – the frame problem, symbol grounding and, of course, the ‘hard’ problem of consciousness. AI need not work thus. In the early days, Turing and Craik sought to focus the mind sciences – not on competencies and tasks – but on ‘intellectual capacities.’ Although we may understand these no better than 50 years ago, we now know that they exploit actions and artefacts as well as the brain’s equivalent of *software*. In 1950, Turing’s imitation game presupposed machines that computed (seemingly) intelligent responses to typed word-strings [46]. Such devices, he thought, could imitate human word-use. This proved to be illusory. Computers calculate better than us, mimic chess-playing and provide practical applications for, say, rudimentary vision. However, they cannot understand word-strings. Without bold thinking, therefore, meaning would never have been traced to self-organizing capacities which integrate neural, bodily and material resources. We would not have realized that, somehow, embodiment grounds semantics.

In the *Nature of Explanation*, Craik focused on our capacity to come up with *objectively valid* models [16]. Meaning depends on *external processes* that prompt us to reason with public symbols. In bridge-building, for example, we agree in our judgements about what symbols mean. Of course, while Craik posited that semantics were exclusively managed by the brain, today we know that this plastic system uses external resources to self-organize [22]. Pure software models are thus giving way to models that leave space for cognitive *deeds*. As this acronym suggests, intelligent activity is dynamic, embodied, embedded, distributed and situated [49]. In popular metaphors, natural born cyborgs use material symbols in a cognitive niche [7, 8]. More sober rhetoric invokes embodied embedded cognition (e.g. [52]) or, perhaps, enactivism (e.g. [43]). Finally, for humans, Hutchins’ work is seminal. To make valid judgements about a

¹ Psychology, University of Hertfordshire, UK; University of KwaZulu-Natal, Durban, South Africa. Email: s.j.cowley@herts.ac.uk

² We explore the *origin of selves* – not at the primordial level of self-other boundaries – but in terms of how we become spinners of narratives [20, 21].

ship's position, naval personnel integrate cultural and physical resources with co-action [29]. Cognition is *culturally distributed*.

Once we look beyond competencies, we discover a domain of culturally distributed meaning. How, then, do symbols guide us to judgements? Instead of positing a language module, we can regard our linguistic skills as the product of self-organizing brains and bodies [22]. Gradually, as we integrate action, perception and external resources (including language), brains set up bio-cultural control systems or *selves*. Ross and Dumouchel say:

Biological systems of the *H. sapiens* variety turn themselves into people—socially embedded teleological selves with narrated biographies in terms of these very beliefs and desires—by taking the intentional stance toward themselves. They can do this thanks to the existence, out in the environment, of public languages that anchor their interpretations to relatively consistent and socially enforced rules of continuity.... [T]hey are incentivized to narrate themselves as coherent and relatively predictable characters, and to care deeply about the dramatic trajectories of these characters they become...[People] are partly constituted out of their social environments, both in the networks of expectations that give identity to them as people, and in the fact that the meanings of their own thoughts are substantially controlled by semantic systems that are collective rather than individual. They are thus not identical to their nervous systems, which are indeed constituted internally. [40] (pp. 264-265)

MacDorman pursues this view by posing the *person problem* [34]. Eschewing the metaphor of embodiment, he asks, "How could human bodies – and perhaps robot bodies –attune to norms and, by so doing, construct themselves into persons?"

3 THE PERSON PROBLEM

Do human bodies use cultural ecology to become persons? To test this hypothesis our focus falls on – not modules – but *cognitive integration*.³ Leaving behind the view that something like software could ever be sufficient to explain intellectual powers, we take a *one system view* [30]. Instead of relying on separating out aspects of nature, we ask how what we do, feel and say arises in bodies-in-the-world. Indeed, as Blackburn suggests, there is no mystery to what is distinct about human agency [5]. To know what someone is doing, we need to see their movements as *expressions of intention and purpose*. However, we lack a viable model of how this is done. Davies writes:

What kind of agent are we? My answer is twofold. First we do not know. At this point in the history of inquiry, traditional notions of agency are dead or dying and their replacements have yet to be born or yet to reach maturity. Second, although we are in a period of conceptual transition and thus we have no developed concept of human agency, we do know what kind of agent we are not." [20] (p. 39)

Once we abandon mentalism, our powers can only be traced to selection mechanisms. Further, given the co-evolution of nature and culture, human forms of agency must be grounded in action and perception. As Davies argues, biological agents use norms –

³ Menary opposes *cognitive integration* to the view that cognitive abilities are solely, or essentially, neural. In minded behaviour, neural and bodily processes are integrated with external vehicles [37,].

nonsymbolic coding – at all functional levels. We must rely on these to become agents who describe our doings in terms of symbolically represented plans and goals. Their basis must arise from "causal or structural capacities which contribute to the exercise of some larger systemic capacity in the larger systemic view [19], p. 4. Nonsymbolic coding, it is argued, operates from the level of organism right down to the cell.⁴

Norm-based biological models provide a basis for exploring how we orient to norms, use them to constrain our dynamics and, by so doing, higher levels of control. They enable us to think about the nature of agency. For this reason the person problem falls in line with the tradition of Craik and Turing. For, in asking how we become persons, we pose a constructive question. Instead of saying why this is difficult, we ask how biology prompts us to use external resources as a basis for developing ways of controlling minded behaviour. Self-organizing processes link cultural resources with the body such that we develop intellectual powers. As a result, we occasionally come up with *objectively valid* judgements. Can robots mimic this? Can they use a changing context to move in ways that humans find appropriate? On this view, flexibility is more important than definition. Indeed, as Wittgenstein saw, even language depends on 'agreement in judgements.' If biological norms allow us to classify particulars (using what Anderson calls *action-guidance representations* [1]), these track aspects of the physical world. Humans, however, also draw on the practices of fellow agents. Given norm-based experience we redeploy our representations as perception becomes associated with how other people are likely to react. Humans thus unite propensities for statistical learning with norm-oriented motivations. Objects, situations and events are constituted, at least in part, by attitudes. Even infants set off contingent responses whose value is monitored and, when rewarded, associated with cues used in anticipatory action [41, 44, 12, 10]. Can robots use human co-ordination in coming to self-construct what (naïve) humans regard as judgements?

4 INTERACTION, FUNCTIONAL-COORDINATION AND CO-ACTION

Although dominant in human life, co-action is not easily characterised. Indeed, in folk psychology, it is often called *interaction*. While readily separated from acting on nonliving things (whose properties produce a nexus of relatively predictable relations), it is more difficult to tell apart from behaviour between living organisms. Recently, interaction has been defined with respect to co-ordination [23]. In encounters between organisms, this comes to constitute an *autonomous* interactional domain. Accordingly, routine co-ordination – when starlings flock or people co-ordinate finger movements – contrasts with *functional co-ordination*. In such cases coupling demands new kinds of integration. Characteristically, this results from disruption. For example, a stickleback may *fail* to perform a mating move or, when walking down a corridor, we may find

⁴ These are 'nonsymbolic' in that they are not reducible to physical tokens that are manipulated in accordance with a syntax; however, in Pattee's sense, they are symbolic [39]. They are rate-independent parameters with a selection history that is inseparable from dynamics. Interestingly, this meshes very closely with Barbieri's view of biosemiosis [4].

ourselves about to collide. In such case, both parties may adjust or co-ordinate functionally. Events beyond the body influence what happens and, in the human case (at least), how each party feels. The achievement of De Jaegher and Di Paolo is to show that this emergent organization can be defined with respect to both mutual influence between individual actions and the concurrent influence of relational dynamics [23].

For Wegner and Sparrow *co-action* occurs when “one agent’s action is influenced by or occurs in the context of another agent’s – and together they do something that is not fully attributable to either one alone” [50]. If we compare this with functional co-ordination, we can be more precise. Whereas De Jaegher and Di Paolo write with biological norms in mind, in human life events are also affected by the cultural nexus [23]. This level of organization enables us to exploit co-action.⁵ Human agents use a history of coping with contingency together with how experience prompts each party to expect the other to react to what is possible. As a result, we develop anticipatory skills by evaluating the ‘something’ that results from co-action. In contrast to association-based learning, this uses – not physical cues – but situation-based evaluations.⁶ Since these set off learning, co-action can channel development. Returning to Wegner and Sparrow, the *something that occurs* underpins agreement in judgements. This unites the child’s intuitive grasp of the value that adults attribute to co-ordinated movements. One party (at least) construes the co-ordinated movements in terms of norms (that can be reported as reasons). Cultural know-how induces them to react to these as expressions of intention and purpose. This attribution has a remarkable outcome. It ensures that, for all parties, a well-timed movement can realize *local values*. Indeed, as Gibson first suggested, this may be the basis of social learning [26, 28].

In illustration, one can consider an event described by Cowley and colleagues [12]. In this, a 14 week old baby interacts with her Zulu speaking caregiver. Strikingly, the baby seems to know that her mother *really* wants silence. When she does so, the reward is a tease and a beaming smile. Co-action arises as the mother captures the baby’s attention with salient, rhythmical in-your-face hand-signals, while using harsh vocalizations to get the baby to behave (these are musical and verbal). Since the event depends on what happens beyond the body, the baby acts neither automatically, imitatively, nor by drawing on any kind of universal. Rather, using experience, both parties orient to the cultural context. To co-act by realising a hoped-for value, the baby uses a pattern that, in Zulu homes, often serves in urging a baby to *thula* (fall silent). Indeed, given local beliefs, caregivers often obtain respect without offering physical rewards (e.g. hugs or physical contact). Where the infant does as hoped for – shows *respect* – this is seen as something special. Acting together with

⁵ In contrast to the approach taken here, it is possible to take the view that there is a continuum from simpler to complex types of functional co-ordination [23]. While this eliminates appeal to intrinsic motive formation (see below), I find it hard to see how it can give an agent the power to realise culturally-based values. This seems to demand the more extended kind of embodiment defended here.

⁶ Arieli & Norton, critique the view that actions are driven by the sum of hedonic aspects [2]. Actions-in-the-world impact on future behavioural trends or ‘contexts are endogenous to decisions’ In this kind of ‘*self-herding*’ past behaviour enables us to derive an arbitrary set point -not necessarily based on hedonic input. If prompted, this functions as input to action and inferences-based-on-action. It thus influences (but does not cause) calculations (and self-reports)

the caregiver, ‘understanding’ is enacted as *both* bodies orient to local norms. The event makes sense because the baby enacts a cultural value (respect). This is more complex than functional co-ordination because the adult habitually realises the value (with variations): she enacts – not movement-types – but a culturally encoded intention (to get her baby to *thula*). Given experience, the baby discovers values realising behaviour. As in this case, this depends on predispositions to use adult response to body-based contingencies. Trevarthen has long argued along these lines [44]. Join affect enables co-action (or *intersubjective behaviour*) to exploit – not just bodily attunement – but neural development based in *intrinsic motive formation* [45, 11].

5 THE TRUTH ABOUT SOCIAL ROBOTS

On the face of things, we might expect social robots to engage in culturally-based co-action or, at least, functional co-ordination. While this may occasionally occur, research has tended to focus on individual competencies. Even social activity is often modelled – not as co-action – but as *imitation* or, at least, as reducible to *syntactic processing*. In such cases, one set of behaviours – or formal patterns – are mapped onto another. In contrast to both functional co-ordination and co-action, such systems fail to go beyond given information or, indeed, cope with even simple disruptions. This logic is defended by Dautenhahn: “Life and intelligence only develop inside a body [which is] adapted to an environment in which an agent lives” [18]. Where the organism is taken to consist in what lies *within* the skin, functional co-ordination seems mysterious. Indeed, models that separate the individual from the world are bound to underplay real-time activity. Autonomy has to be conceptualised in terms of individuals who are “embedded, coupled and linked to a social context” [18]. Placing the living on the inside, context is idealised around external invariants that, it is hoped, can be discovered by *complete systems* [18]. Up to the present, then, social robotics has adopted what Järvillehto decries as the two-systems view [30]. For example, in Breazeal’s work, what matters is mainly whether the robot supports the model that the engineer has assigned (if it displays a competency) and secondarily whether this can be successfully displayed in the ‘assigned interaction scenario’ [6]. By contrast, if the kinds of co-ordination displayed by the living are the basis for social life, we are in for a surprise.

Emotion-showing robots such as Kismet are not social. They manifestly fail to display the adaptive, flexible behaviour that is typical of living systems (even at the level of a single cell). This is because they have been designed – not with an eye to biology – but to replicate competencies. Building on folk views of the individual, they are expected to contribute to the doings of heterogeneous groups of humans and/or other robots. The putative competencies are expected to scale up in ways that will eventually give us ‘embodied agents’ who perceive and interpret the world in ways that enable them to ‘recognise each other and engage in social interaction’ [18]. Ideally social robots will become individualised agents who, ‘on the basis of their own experience’ can ‘explicitly communicate and learn from each other’ [18]. Since they – and, by implication, we – only simulate sociality, the approach introduces many puzzles. For one thing, interaction seems to be an inadequate basis for explaining either sentience or social perception. With current

technologies, machines lack the sentience that shapes individual experience and, without this, it is hard to see how they could acquire the ‘symbols’ needed in explicit communication and learning. Indeed, one runs into all the problems that arise from reducing human powers to the competence models that are the mark of mentalist tradition.

The person problem presents social robots as part of a single system or cultural ecology that includes humans. Instead of highlighting individuals, the question becomes how they can be integrated into our cultural world by using co-ordination and, ideally, developing a manifest sensitivity to social norms. To do this, of course, social robots will have to mimic bio-cultural aspects of agency. Using cognitive integration, like human infants, they need to detect cultural norms that can be used to reshape their control systems. Given what we know of infants, this depends on physical objects, interactional relations the evaluation of (joint actions) and how contingencies map onto local norms. While lacking space to discuss these matters, today’s synthetic models already use such resources. For current purposes, however, I highlight another issue raised by the person problem. To understand our ecology, we can study robots *in the wild*. As Hutchins did in studying naval practice, we can ask how agents use social robots as resources that, among other things, lead them to agreement in judgements [38]. Thus, rather examine how robots influence us. Just as was done for computers, we can take on board the idea that a primary use of social robots is, at this stage, as tools that can be used to generate and test hypotheses about minded behaviour.

While not designed to use co-ordination and social norms, we *feel* that social robots are social. Indeed, they are designed to make us treat them more like living dolls than fancy computers. As is widely attested, even simple robots become *relational objects* [47] that set off anthropomorphic reactions. Larger, more robust machines can have a dramatic impact on a cultural ecology. This appears dramatically when Robovie, a Japanese-designed social robot, is introduced to classroom environments. Even if the machine’s interactional routines are governed by software, humans attempt to set off co-action [42, 38]. In longitudinal trials, the classroom ecology alters in dramatic ways. Further, while initial interest may fall off, the robot evokes high quality interactions and, for some of the children at least, these are maintained over time. The outcomes have been extensively studied: work includes several case studies, the micro-coding of response to robot behaviours, attention to global variables labelled sociability and familiarity and, recently in *content behavioural analysis* [15]. In this latter approach, the focus falls directly on content-patterns that arise as children structure their co-action around the machine. Thus, as in traditional content analysis, human-human-robot analysis is described in terms of recurring themes.

6 LEARNING FROM HUMAN CO-ACTION

Robovie functions as a social mediator. While long recognised that robots *mediate contact between children* [51], the idea of a social mediator has a long historical tradition [38]. Using micro-analysis, Nabe and his colleagues show that robot behaviours can take on language-like meaning or, in Vygotsky’s (1981)

terms, become *mediational means* [48].⁷ They function as psychological tools by allowing human agents to access a nexus of cultural norms. Indeed, as Hegel and Marx argued, such means locate the subject in a historical context. In the classroom, mediational means function to spur study and play that enable children to develop both individual skills and co-action routines. While many robot behaviours elicit, at best, stock responses, others become increasingly valued [38]. In short, while most seem pointless (and are ignored), others set off enjoyable events. This is not surprising. Babies too ignore most of what happens but, given certain sorts of attention, make co-action attempts. Unlike baboons (but perhaps not chimpanzees) much of our behaviour is directed –not at goals –but through experience of external resources. Humans act epistemically [14]. We act in world-directed ways by seeking out interesting effects. In the classroom, this happens because children anthropomorphise the robot by seeking to set off social action. Indeed, were the robot human, what they do would induce rewarding response based on shared orientation to local norms. The robot, however, does not perceive fine human movements – let alone their social meaning. For this reason, it is incapable of either functional co-ordination or, by extension, evaluation of co-action attempts. Given the classroom ecology, however, children respond to *each other*. Thus certain robot-directed co-action attempts become valued [29]. Ways of orienting to robot behaviours become contexts that can be used to set up achieving social effects – one’s based on co-action. Using a semantic nexus associated with a movement, children learn from robots.

Can robots learn from humans? Of course. However, machine learning tends to be conceptualised in terms of specific tasks. Typically, what is modelled presupposes a competence that is the theoretical descendent of concepts like the *mental lexicon*. Then, using this model, the system learns tasks such as associating sound-patterns (words/phones) with images or acoustic patterns. To address the person problem, however, human agency can be traced to how, in real-time, we orient to social norms. Instead of evoking competencies or minds, we can focus on how acting jointly, we come to realise values. Are words be learned this way? Elsewhere Cowley presents this as the basis for *human symbol grounding* [10]. This, then, is a separate issue from learning to use physical invariances to track salient aspects of the world. In other words, as developmentalists often argue (e.g. [32]) rather than relying on linking mind to world, we rely on linking what is perceived to expectations based on the experience of persons. This leads to simplification. Above all, objects can be recognised with respect – not to invariances – but culturally salient aspects. Thus babies come to perceive, say, moods, possible actions and the likely arrival of dinner. To simulate functional co-ordination, therefore, it would be possible to rely exclusively on programming. Robovie, for example, might use sensor input to calculate the laughter-frequency of individuals and, using rules, produce canned laughing (or other behaviour) to join in with them. There is little doubt that this kind of integrational process would have a massive impact on the classroom ecology.⁸ While there is

⁷ Vygotsky listed counting, mnemonic techniques, algebraic symbol systems, works of art, schemes, diagrams, maps, mechanical drawings, all sorts of conventional signs and so on [48] p. 137.

⁸ Robovie uses pseudo learning. After a certain number of interactions with the device a child is rewarded with a secret. This is an effective technique that has striking classroom effects [13].

nothing intelligent about this, much the same goes for human laughter. This too has an automatic basis. What is striking, then, is how we are able to draw on a cultural nexus in learning to control when – and how – to laugh. That, however, is complex: it is already a clear case of co-action.

Programming could also be used to simulate co-action. One could hard-wire biases that prompted the robot to identify a subset of behaviours. Second, with more hard-wiring, these could be mapped onto cue-defined context-types. In the laughter example, the robot could distinguish this from other vocalizations and relate both to the t-behaviour (teasing) that recur in the Japanese classroom [15].⁹ Using cues like whether, at time *t*, more than one child laughs – or all laugh – follow-up response could be varied. Third, response-by-laughing could, as it is in the children, exploit capacities for synchrony, simultaneity and setting up sequences. This would change their encounters. However, less dramatic examples might have more powerful effects. Given the (rather stupid) ways in which children rub and tap different sensitive parts of the robot, interesting simulations of co-action could be induced by timing vocal patterns that simulated affect which were systematically associated with kinds of touching. There is little doubt that children would be enthusiastic. Contingencies and co-ordination are at the heart of human social life. In proposing this approach to social robotics, therefore, the focus falls on using robots as test-beds to how, given these physical resources in this culture, humans construct co-action.

7 ADDRESSING THE PERSON PROBLEM

Could design replace simulation? Might machines use movement to detect affect and simulate the motive formation which underpins co-action? If so, we might make progress with the person problem. The difficulty, if there is one, lies in identifying – not physical patterns – but variable contingent cues associated with abstract norms. A machine would have to *discover* something as does the Zulu baby. If control-side, the agent is in state *X*, it must calculate that world-side cues *mean* either (a) inhibit; or (b) do not inhibit. Depending on a decision, a baby is likely to go into either state *Y* (e.g. hurting less) or state *Z* (e.g. crying more). Finally, this will influence longer-term rewards (maybe being abandoned; maybe being picked up). Of course, such decision making does not develop from scratch: like the baby, the robot would use experience to form motives while using biased sensors to predict likely up-coming rewards. This may be difficult. Nonetheless, a machine that did this would meet several of the proposed benchmarks: it would function as an embodied agent that simulated *experience to perceive and interpret* a small aspect of *the world*. While this would not be sufficient for social perception or the development of capacities for fully-fledged interaction, such an agent would learn from how humans co-act. Crucially, it would act – and not act – by integrating its own states with human propensities to reward certain norm-based co-actions. It would function as part of the cultural ecology – as a one-system model or an an integrated mediational means.

Much empirical work points towards ways of approaching the person problem. What is more challenging is

⁹ In the current model of content-behavioural analysis, three of the fifteen themes that can be associated with teasing.

that the approach demands ways of integrating research based in, for example, simulations, ALife and robotics. New kinds of collaboration are required. By way of illustration, let us sketch recent progress in ALife and Android Science. Thus even artificial environments can be used to detect variable contingent cues that are needed for functional co-ordination [3]. While independent of historically-derived norms, the world serves in shaping an agent's powers. Given central nervous systems (or computation), variable contingent cues can take on many functions.¹⁰ To develop into simulations of co-action, different sensorimotor means would be needed in integrating activity. Here android science provides machines that not only elicit culturally appropriate response, but rely on a human subject's (norm-based) expectations. For example, gaze-behaviour is human-like if – and only if – people believe that the android is under human control [35]. Gaze is a co-actional resource: looks function in the context of another agent's gazing – and, given a nexus of norms, set off something which can be attributed to neither party. In human life, this strange property is central to both mother-infant interaction and, say, flirting. Going beyond the information given is equally compatible with the emerging view that language – far from depending on symbolic competencies – derives from the physics of expressive co-action. Androids, of course, are the ideal test-bed for exploring how expression is integrated across modalities and between parties [33]. Even today, they could be programmed to prompt humans to use gaze in jumping to conclusions that are, in many cases, unwarranted¹¹. It is an empirical question whether they could discover how gaze contributes to co-action. Strikingly, roboticists may be in a strong position to give new insight into how gaze is integrated with both other expressive dynamics and verbal patterns.

8 TOWARDS CO-ACTING ROBOTS

It is so hard to overthrow mentalism that even in the embodied, embedded tradition some invoke extended *mind* [9]. As Gibson suggests, this may be because human learning relies on meaning and value [26, 28]. Indeed, in an early paper, he argues that a viable learning theory will draw on – not behaviourism – but social psychology. Only then can learning be seen, not as the basis for social life but, rather its consequence. As emphasised above, humans act to realize values. In focusing on co-action, we propose that human bodies construct themselves into persons by using the movements of others. The trick, however, lies in being designed to *take these* as expressions of intention and purpose. For this to be possible, human babies exploit *intrinsic motive formation* [45]. As a result they can use not only disruptions that lead to functional co-ordination but also cues that mark historically derived cultural norms. Gradually, they become biocultural agents. We construct ourselves into persons by learning how to use the doings of other in realizing values.

It has not been demonstrated that co-action cannot derive from functional co-ordination. *That* is an empirical question. However, if functional co-ordination is sufficient for values realizing behaviour, even intellect can be traced to embodiment.

¹⁰ This could be construed as a way of talking about affordances.

¹¹ While extreme gaze sensitivity is especially prominent in psychotics, it is rare only among people in the autistic spectrum [17].

On the view presented here, however, co-action introduces a new level of complexity. Just as functional co-ordination depends on mutual influence between individual actions and the concurrent influence of relational dynamics, co-action exploits norms that are realised by material acts that carry historical patterns. In this way real-time relations can be linked with the slow dynamics associated with habits. Values can be realised by orienting to a cultural nexus of norms or the *interaction order* [27]. Slowly we develop shared models of interaction-types. By the age of 4 or 5 children get a sense of what, for example, distinguishes a doctor from a patient. Cultural norms associated with verbal patterns prompt thoughts about which people agree. Given their growing sense of what is expected, children come to act strategically. With agreement in judgements, their views come to have some objective validity – “Doctors make you stick your tongue out.” Children become role-playing agents whose preferences draw, in part, on experience and, in part, on the rewards of self-display. In learning about strategic signalling, we discover how to play and, gradually, turn into persons. We become skilled with a range of practices and, in the end, may learn the tricks and practices that, as in Craik’s [16] example, enable us to interpret symbols as the plans for a bridge. This depends, not on inner competencies, but discovering the values needed by a bridge-building engineer. By posing the person problem, we suggest that much can be gained by tracing social skills to co-action: this, it is hypothesised, is nature’s trick for getting us to treat visible and vocal gestures (speech) as expressions of intention and purpose.

REFERENCES

- [1] Anderson, M. & Rosenberg, G. Content and Action: The Guidance Theory of Representation. In: *Evolutionary Biology and the Central Problems of Cognitive Science*, D. Smith (Ed.) *Journal of Mind and Behavior*, 2007.
- [2] Ariely, D. & Norton, M. How Actions Create -not just Reveal Preferences, *Trends in Cognitive Science*, 12/1: 13-17. (2007)
- [3] Auvray, M., Lenay, C. & Stewart, J.. The Attribution of Intentionality in a Simulated Environment: the Case of Minimalist Devices. In: *Tenth Meeting of the Association for the Study of Consciousness*. Oxford, UK, 23-26 June, (2006).
- [4] Barbieri, M. Is the Cell a Semiotic System? In: *Introduction to Biosemiosis: the New Biological Synthesis*. Dordrecht: Springer, M. Barbieri (Ed.) pp. 179-208. (2007).
- [5] Blackburn, S. *Ruling Passions*, Oxford: Clarendon Press. (1999).
- [6] Breazeal, C. Towards Sociable Robots. *Robotics and Autonomous Systems*, 42/3-4: 161-175. (2003).
- [7] Clark, A. *Natural Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford: Oxford University Press. (2003).
- [8] Clark, A. Language, Embodiment and the Cognitive Niche. *Trends in Cognitive Sciences*, 10/8: 370-374. (2006).
- [9] Clark, A & Chalmers, D. The Extended Mind. *Analysis* 58(1): 7-19. (1998).
- [10] Cowley, S.J. How Human Infants Deal with Symbol Grounding. *Interaction Studies*, 8.1: 83-104.. (2007).
- [11] Cowley, S. J. The Codes of Language: Turtles All the Way Up? In *The Codes of Life*, M. Barbieri (Ed.) 319-345. Springer, Berlin. 2007.
- [12] Cowley, S.J., Moodley, S. & Fiori-Cowley, A. Grounding Signs of Culture: Primary Intersubjectivity in Social Semiosis. *Mind, Culture and Activity*, 11/2: 109-132. (2004).
- [13] Cowley, S.J & Kanda, T. Friendly Machines: Interaction-Oriented Robots Today and Tomorrow. *Alternation*, 12.1a: 79-106. (2005).
- [14] Cowley, S.J. & MacDorman, K.F. What Baboons, Babies and Tetris players tell us about Interaction: a Biosocial view of Norm-based Social Learning. *Connection Science*, 18/3, 363-378. (2006).
- [15] Cowley, S.J., Langford, D. & Schulz, J. Content coding: tracking a social mediator’s achievements. Paper to be presented at *Human-Robot Interaction*, Amsterdam (2008).
- [16] Craik, K. *The Nature of Explanation*. Cambridge: Cambridge University Press. (1943).
- [17] Crespi, B. & Badcock, C. (in press). Psychosis and autism as Diametrical Disorders of the Social Brain. *Behavioral and Brain Sciences* (to appear).
- [18] Dautenhahn, K. Embodiment and Interaction in Socially Intelligent Life-like Agents. In C. Nehaniv (ed) *Computation for Metaphors, Analogy and Agents*. Springer Lecture Notes in Artificial Intelligence, Volume 1562, pp. 102-142. (1999).
- [19] Davies, P. S. *Norms of Nature: Naturalism and the Nature of Functions*, Cambridge MA: MIT Press. (2001).
- [20] Davies, P. S. What kind of Agent are we? A Naturalistic Framework for the study of Human Agency. In: *Distributed Cognition and the Will*, D. Ross, D. Spurrett, H. Kinkaid & L.G. Stephens (Eds.) MIT Press: Cambridge MA, pp.39-60. (2007).
- [21] Dennett, D. C. The Origins of Selves. *Cogito*, 3, 163-173. (1989).
- [22] Dennett, D. C. *Consciousness Explained*, Boston: Little, Brown & Company. (1991).
- [23] De Jaegher, H & Di Paolo, E. Participatory Sense-making: An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485-507. (2007).
- [24] Di Paolo, E., Rohde, M., Izuka, H. Sensitivity to Social Contingency or Stability of Interaction? Modelling the Dynamics of Perceptual Crossing. To appear, *New Ideas in Psychology*. (2008)
- [25] Fong, T., Nourbakhsh I., & Dautenhahn, K. A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems* 42(3-4), 143-166. (2003).
- [26] Gibson, J. J. The Implications of Learning Theory for Social Psychology. In: *Experiments in Social Process: A Symposium on Social Psychology*, J. G. Miller (Ed.) McGraw-Hill: New York, pp. 149-167. (1950).
- [27] Goffman, E. The Interaction Order, *American Sociological Review*, 48, pp. 1-17. (1983).
- [28] Hodges, B. Good Prospects: Ecological and Social Perspectives on Conforming, Creating and Caring in Conversation. *Language Sciences*, 29: 584-604. (2007).
- [29] Hutchins, E. *Cognition in the Wild*. Cambridge, MA: MIT Press. . (1995).
- [30] Järvillehto, T. The Theory of the Organism-Environment System 1. Description of the Theory. *Integrative Behavioural and Physiological Science*, 33/4: 321-344. (1998).
- [31] Kanda, T. Sato, R. Saiwaki, N. & Ishiguro, H. A Two-month field Trial in an Elementary School for Long-term Human-robot Interaction. *IEEE Transactions on Robotics (Special Issue on Human-Robot Interaction)*, 23(5), pp. 962-971. (2007).
- [32] Legerstee, M. *Infants’ sense of People: Precursors to a Theory of Mind*. Cambridge University Press, Cambridge.. (2005).
- [33] MacDorman, K. F. & Ishiguro, H. “The Uncanny Advantage of using Androids in Social and Cognitive Science Research”, *Interaction Studies*, 7(3) 143-158. (2006).
- [34] MacDorman, K. Life after the Symbol-system Metaphor. *Interaction Studies* 8.1: 143-158. (2007).
- [35] MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S. J. & Ishiguro, H. Assessing Human Likeness by Eye Contact in an Android Testbed. *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society*. (2005).
- [36] Menary, R. Writing as Thinking. *Language Sciences*, 29; 621-633. (2007).

- [37] Menary, R. *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan. (2007).
- [38] Nabe, S., Cowley, S.J., Kanda, T. Ishiguro, H. Iraki, K. & Nargita, N. Robots and Social Mediators: Coding for Engineers. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication. University of Hertfordshire*, pp. 384-390. (2006).
- [39] Pattee, H. The Physics of Symbols: Bridging the Epistemic Cut. *Biosystems*, 60: 5-21. (2001).
- [40] Ross, D. & Dumouchel, P. Emotions as Strategic Signals. *Rationality and Society*, 16(3), 251-286. (2004).
- [41] Sommerhoff, G. & MacDorman, K. F. (1994). An Account of Consciousness in Physical and Functional terms: A Target for Research in the Neurosciences. *Integrative Physiological and Behavioral Science*, 29(2), 151-181.
- [42] Tanaka, F. Cicourel, A. & Movellan, J.R. Socialization between Toddlers and Robots at an Early Childhood Education Center. *Proceedings of the National Association of Sciences*. (to appear).
- [43] Thompson, E. *Mind and Life: Biology, Phenomenology and the Sciences of Mind*. Cambridge MA: Harvard University Press. (2007).
- [44] Trevarthen, C. The Concept and Foundations of infant Intersubjectivity. In: *Intersubjective Communication in Early Ontogeny*. Cambridge: Cambridge University Press S. Bråten (Ed.), pp. 15-46. . (1998).
- [45] Trevarthen, C. & Aitken, K. J. Infant Intersubjectivity: Research, Theory and Clinical Applications. *Journal of Child Psychology and Psychiatry*, 42(1), 3-48. (2001).
- [46] Turing, A.M. Computing Machinery and Intelligence", *Mind*, 49, pp. 433-460. (1950)."
- [47] Turkle, S., Taggart, W., Kidd, K. and Daste, C. Relational Artifacts with Children and Elders: the Complexities of Cybercompanionship. *Connection Science*, 18/4: 347-362. (2006).
- [48] Vygotsky, L. The Instrumental Method in Psychology. In: *The Concept of Activity in Soviet Psychology*, J.V. Werstch (Ed.) Armonk, NY: N.E. Sharpe, pp.134-143. (1981).
- [49] Walmsley J. Methodological Situatedness; or, DEEDS worth Doing and Pursuing. *Cognitive Systems Research*, (2008).
- [50] Wenger, D. & Sparrow, B. The Puzzle of Coaction. In: *Distributed Cognition and the Will*, D. Ross, D. Spurrett, H. Kinkaid & L.G. Stephens (Eds.). MIT Press: Cambridge MA, pp.17-41. (2007).
- [51] Werry, I., Dautenhahn, K. Ogden, B. Harwin, W. Can Social Interaction Skills be Taught be a Social Agent? The Role of Robotic Mediator in Autism Therapy. Springer Verlag, *Lecture Notes in Computer Science, subseries Lecture Notes in Artificial Intelligence*. (2001).
- [52] Wheeler, M., *Reconstructing the Cognitive World: The Next Step*. MIT Press, Cambridge MA. (2005).