Extending Trust: Coupled Systems, Trust and the Extended Mind

Neal R.T. Leblanc (nleblanc@silverblaze.net)
Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive
Ottawa, ON, K1S 5B6 Canada

Abstract

In this paper, I attempt to to examine the concept of reliability in Extended Cognition, using frameworks and data from social and evolutionary psychology to examine two of the criteria: transparency and endorsement. Using this framework, I will argue that the seemingly contradictory experimental results in Extended Cognition research are the result of ignoring the differences between types of cognitive artefacts (active vs. passive) and the higher levels of trust required for active artefacts to be considered reliable as a result of our ascribing them agency.

Keywords: Extended Cognition; Epistemic Structures; Trust; Distributed Cognition; Agency.

Introduction

In their seminal paper "The Extended Mind," Clark and Chalmers (1998) put forward what appears to be a somewhat radical claim: that cognition is not bound within the confines of the skin and skull. They argue that making use of cognitive technologies as part of the cognitive process produces a powerful, two-way interaction between the human and the artefact. This interaction results in a coupled system, such that "all components play an active and causal role, and they jointly govern behaviour" (1998). As a result of the complex and non-linear interaction, the performance and ability of the system as a whole is greater than and cannot simply be explained as the simple sum of the capabilities of its components. Further, removing any component of the coupled system (be it the tool or a neural cluster) will cause an overall reduction in the system's competence. Thus, rather than arbitrarily using the skin as a barrier to determine what is part of the cognitive system, they argue that *reliability* should be the salient discriminatory characteristic for what is part of the cognitive system and what is not. They argue that reliability consists of three criteria: *availability*, *transparency* (automaticity of use) and *endorsement* of the artefact and its content (trust).

The purpose of this talk is to examine the concept of reliability in Extended Cognition, using frameworks and data from social and evolutionary psychology to examine the individual criteria (save availability, which is remarkably straightforward). Building on this discussion, I will attempt to reconcile seemingly contradictory experimental results in Extended Cognition research.

Simple Cognitive Artefacts

The simplest and most common type of cognitive artefact is the epistemic structure (or artefact): a construct made in the environment which serves to hold information. The ability to create such structures has evolved in many species, both complex and simple, across the spectra of nature. Many insect species (including ants and termites) employ pheromone trails to allow them to easily return to a food source or to warn against danger, such as predators (Camazine et al., 2001). Higher order animals frequently use scents or visual aids to mark trails within their territories, as well as to lead them back to caches of food they have made (Sterelny, 2004). While these examples many not be as elegant or representationally rich as humanity's written words, they serve the same purpose. They allow for the offloading information from the organism into the environment. Thus, as Dennett (1996) says: "This widespread practice of off-loading releases us from the limitations of our animal brains."

It is important to note, however, that despite the complexities of language and numerals, even the epistemic structures used by humans are often quite simple. In fact, these simpler epistemic structures are "everywhere" in human life (Kirsh, 2006). Many of these structures simply serve to ease our memory burden, as when we keep our keys near the door or put something that needs to be mailed under our keys. However, we also alter our environment in order to convert complex tasks into simpler ones. Examples of this are legion, ranging from the simple act of marking a trail to simplify later navigation, to the organization of important notes into a filing system, to the complex behaviour of skilled bartenders who use both the sequence and shape of bar glasses in order to optimize their performance (Clark, 2001a). Kirsh & Magilo (1994) call such environment-altering behaviour *epistemic actions*. Using a simple Tetris-player task, they demonstrate not only that people perform such actions (despite being literally counter-productive in terms of purely pragmatic game-play efficiency), but that the number of such actions taken was strongly predictive of task performance. In a follow-up longitudinal study (Magilo & Kirsh, 1996), it was shown that the number of such epistemic actions (and associated backtracking) increased with the skill level of individual participants, indicating that it was an effective learned strategy. Thus, as Clark (1997) put it, "We use intelligence to structure our environment so we can succeed with less intelligence. Our brains make the world smart so we can be dumb in peace!" (p. 180).

After more than a decade of study, Kirsh (2006) notes that such structures and actions are generated so simply and automatically that they often go unnoticed by both researchers and the people making use of them. As a result, he suggests that perhaps the only way to study them is to record a person's behaviour, and then perform an ethnographic analysis after the fact.

Cognitive Artefacts in a Shared Environment

Sterelny (2004) argues that Clark (and, by extension, other proponents of extended cognition) has made a critical error in his picture of the extended mind and of epistemic artefacts. Specifically, Clark focusses only on tools being used by a single agent, whereas offloaded epistemic structures exist in the shared environment and are often themselves shared, and are thus subject to interference. Sterelny provides a detailed evolutionary account of the use of tools and epistemic artefacts, which stresses the importance of the evolution and use of social guards (tricks which we employ in order to protect and validate the data in the environment and to detect cheating by members of our social group), especially in light of evolutionary pressure to get a free ride by making use of the epistemic structures of others (or, for that matter, manipulating the structures of competitors or prey). He contrasts these with purely internal resources that are not exposed to outside manipulation, and thus do not need to be vetted. Sterelny believes that cheater detection is "a problem whose informational load is both heavy and unpredictable" (2004), and therefore argues that, as a result, we have a tension between two of the criteria of reliability: *transparency* and *endorsement*. The deployment of social guards when dealing with external resources generates high demands on our cognitive economy, increasing attention and processing, thus endangering the automatic endorsement which is required for an external resource to count as part of the mind. Thus, in order to endorse the content of something, its use is no longer automatic. Sterelny takes this even further, arguing that the cognitive costs of coupling are higher than the benefits that would be gained. Sterelny (2005) does, however, allow that some social guards may themselves be offloaded into the environment (such as our ability to recognize our own handwriting).

In response, Parsell (2006) argues that Sterelny is likely overestimating the cost of the use of social guards. First, Parsell demonstrates that a simple connectionist network can be created which performs cheater-detection without requiring any additional modules, thus showing that the processing costs of some types of cheater-detection may be trivial. Furthermore, following Sterelny's admission that the social guard task may itself be partially offloaded, Parsell discusses the use of passwords in modern technology, and perhaps more importantly, makes a case that the perception of continued possession of an artefact creates a (possibly misplaced) strong endorsement of its contents, seemingly bypassing or negating the need for social guards.

Chandrasekharan & Stewart (2007) use an evolutionary computer model to demonstrate that strategies for use of epistemic structures can occur as a result of evolutionary pressures, at least in synthetic agents. Further, they demonstrate that the use of epistemic structures not only lowers cognitive load (countering Sterelny's concerns about the cost being too high), but postulate that this lowering could, in fact, drive the generation of additional structures, essentially leading to bootstrapping.

Transparency and the Costs of Coupling

Despite the fact that Sterelny's hypothesized expensive social guards do not appear to be present, the use of cognitive artefacts is not completely without cost. The communication link between agent and artefact is itself an information processing task which involves the encoding and decoding of information and the activation of the perceptual system, at the very least. The act of activating the coupling link, however, appears to be nearly automatic:

"Biological brains ... are by nature open-ended controllers. To deal fluently with bodily change and growth, they have developed ways of computing, pretty much on a moment-to-moment basis, what resources are readily available and under direct control" (Clark, 2005).

The decision to couple or not is determined by a quick cost-benefit analysis of the perceived utility of the artefact against the cost of its use, evaluated on a case-by-case basis (Lee & Moray, 1992; 1994). This analysis, however, seems to be unbiassed in its selection of which resources to apply to a given problem, be they external or internal. Gray et al. (2004; 2006) demonstrated this experimentally by having subjects perform a task with the option of using an automated assistant during a simple cognitive task: programming a simulated VCR. They conclude that the "control system is indifferent to the information source" (2006). What is important is the cost of using the aid, which they conclude is simply a function of reaction time, at least for this non-critical task. These data seem intuitive, if one considers the task of adding two single-digit numbers. In such a case, the perceived utility of the calculator is so small that even if one is close at hand, it is only rarely used - whereas people will expend large amounts of effort and energy to find a calculator when faced with more complicated mathematical tasks.

While *activating* a coupling link appears to be an automatic task, building that link initially is itself a learned behaviour. There is a cost in time and cognitive resources that must be paid in order to integrate a new and novel artefact into our cognitive systems (Karwowski, 2000). Furthermore, the cost of integration is not fixed, but depends on the complexity of the artefact. This is the basis of Karwowski's Complexity-Incompatibility Principle: "As the artifact-human [sic] system complexity increases, the compatibility between system elements, expressed through their ergonomic interactions at all levels, decreases, leading to greater ergonomic entropy of the system." (2000) Thus, he argues that special care must be taken in the design of artefacts in order to assure compatibility with humans.

Sutton (2006) presents a similar view, arguing that much of modern human cognition is a result of what he refers to as the "soft assembly" of transient and repeatable systems involving both internal and external representations and resources. As a result, our neural resources come to be "expressly tailored to accommodate and exploit the

additional representational and computations potentials introduced" (2006) as we integrate those devices which we find to be useful. This is reinforced by research which shows that our plastic minds incorporate tools into the body map, and become accustomed to and anticipate the feedback these tools provide (Hawkins, 2004). Thus, true coupling occurs when we go beyond the "soft assembly" by integrating an artefact which we have found to be highly reliable and either highly durable or frequently available, such that the "new capacities are sufficiently robust and enduring as to contribute to the persisting cognitive profile of a specific individual" (Sutton, 2006).

Trust and Complex Artefacts

Perceived utility of an artefact is an especially strong concept amongst researchers in human-computer interaction and the psychology of trust in automation, where it serves as the core component (if not the very definition) of trust in technology and automation in many frameworks (e.g. Lee & Moray, 1992; 1994; Fogg & Tseng, 1999; Dzindolet, Pierce, Beck & Dawe, 2002; Riegelsberger, Sasse & McCarthy, 2005; Kaasinen, 2005). In this literature, the perceived utility is defined as the comparison of the user's assessment of the system's performance versus the user's assessment of their own performance, an analysis which is highly subject to bias.

According to the trust framework put forward by Dzindolet, Pierce, Beck & Dawe (2002), disuse of a cognitive artefact (which is to say, choosing not to use the artefact even when it would be appropriate) is a result of mistrust of the artefact. In their experiments, they demonstrate that users have an initial expectation of computer superiority (which they call the *automation bias*), however, errors made by the systems are extremely salient. Specifically, people rated performance of automated systems as lower than their own, even when the system made less than half as many errors and non-cumulative feedback was provided at each trial. By contrast, subjects were more lenient and trusting of "human experts" with the same or worse performance profiles as the automated system. They assert that the automated system is betraying our initial trust by making its errors (thus violating our expectations), and thus is quickly judged to be untrustworthy. One simple example that they present is the case of automated alarm systems, and what has come to be called the *cry wolf* effect. Essentially, only a few false alarms signals are required to greatly degrade trust in the system, and, accordingly, response to the alarm. Dzindolet et. al also demonstrate the rapid-distrust effect experimentally. In this study, subjects were asked to perform a task, and after each of their responses, they were shown the response of an "aide," which was either described as a human expert or a computer program. At the end of the study, subjects were offered a reward based on the accuracy of a randomly selected sample of the answers from the previous trial, and allowed to have the reward calculated based on their own answers or that of the aide. Surprisingly, even when told explicitly that the automated system made less than half as many errors as they did, 81.25% of subjects chose to use a selection of their own responses rather than those of the automated responses. By comparison, when told that the automated responses were actually those of a human expert, 50% of subjects chose to use the judgement of the aid. Thus, they conclude that people interact with machines somewhat differently than they do with humans.

Artefacts and Agency

These results stand in contrast with those of Reeves and Nass (1996), who demonstrate that humans exhibit behaviours with computers similar to their behaviours with other humans. Such examples include attraction to agents whose characteristics are most like their own, a greater willingness to accept flattery than criticism from the computer, and a less critical approach to the computer directly, rather than "behind its back." Based on these results, Reeves and Nass conclude that human-computer interaction is natural and social in nature.

Miller (2004) uses these results to argue that computers have bypassed what he refers to as the "agentification barrier," a point where an artefact reaches a sufficient level of complexity and autonomy that we ascribe qualities such as intent and awareness to it. Miller demonstrates that this difference is so pronounced that we even use different language when referring to computers rather than other tools: "Even my language, as I write this, is illustrative: I hit myself with the hammer, while my computer does things to me" (2004). As a result, Miller claims that humans readily generalize their expectations from human-human interaction to human-computer interaction regardless of whether or not that is the intent of system designers.

Framing Trust

Riegelsberger, Sasse & McCarthy (2005) follow a similar track as they lay out what they believe to be a *general* framework for trust, encompassing both human-human interaction and human-computer interaction. They claim that the largest difference between trust in technology and trust in humans is that, when dealing with automation, the primary issue is the trustor's perception of the trustee's ability, since computers do not have motivation. However, they contend that most technological "agents" are in fact part of a larger socio-technological system and should thus be analysed using the entire framework. This is especially true, they argue, due to the fact that some users are likely to *ascribe* motivation to the automated agent (as per Miller's agentification). They assert that simple social guards are in a constant state of evolution as the guidelines for what should make an agent trustworthy are co-opted by untrustworthy actors, thus reducing or eliminating their value, as can be shown with the increased complexity of internet phishing scams. Simply increasing the number of social guards is also not a viable option, because then the burden of trust-testing takes over the entire transaction,

causing Sterelny's argument that the cost of use outweighs the usefulness to materialize. Thus, they argue, while trustworthiness "markers" do contribute to *perceived* trustworthiness, they are not by themselves sufficient to generate trust.

Thus, beyond simple markers, a trustor must rely on cues from the trustee and the environment in order to assess both the ability and the motivation of the trustee. From this, five factors of trust are posited, and are split into *external* and *intrinsic* groups (Riegelsberger, Sasse & McCarthy, 2005). External factors are pressures which act to coerce the trustee into compliance. These include *temporal embeddedness*, or the prospect of later retaliation; *social embeddedness*, or the prospect of the trustee's reputation; and *institutional embeddedness*, which is a combination of the trust in the brand associated with the trustee and the trust in the society which creates regulations to which they must conform or risk punishment. The intrinsic factors are: *ability*, which is the belief that the trustee is able to perform the task (*perceived utility*); and *internalized norms*, which, in the case of technology means dependability – that the system will continue to work in the same way over time. Since the external measures of trust are used to measure the motivation of a trustee, they are less salient when evaluating the trustworthiness of an automated agent. In fact, it is unclear that an automated system is embedded either temporally or socially. Institutional embeddedness, however, does appear to be a factor; sociologists are showing that everyday interactions are increasingly based on trust in a brand rather than the individual (Riegelsberger, Sasse & McCarthy, 2005), which creates an obvious extension to computer-based agents, especially when used for commerce.

Affective Trust

Riegelsberger, Sasse & McCarthy (2005) state that the lack of trust in technology can be partially attributed to a lack of interpersonal cues. Citing research by Rickenburg & Reeves (2000), they show that some cues lead to an affective trust even if there is no rational reason for this trust. For example, the use of a synthetic voice or a synthetic animated character with only very basic interpersonal cues was found to increase trust.

Schaumburg (2001), on the other hand, argues that trustworthiness does not come as a consequence of painting a face onto an agent's interface. In fact, he makes the claim that in some cases, such an interface may increase the user's anxiety rather than decreasing it, depending on the nature of the social interaction and whether or not the user initially overestimates the agent's usefulness. His claim is based in part on a study by Van Mulken, André and Müller (1999) in which users did not follow the recommendations of an anthropomorphic agent (such as a cartoon character) more readily than a non-anthropomorphic one (such as a text or audio message), and did not rate it as any more trustworthy. It would be interesting to determine if these data differ due to purely methodological reasons (since, for instance, audio messages were considered anthropomorphic in one study but not the other), or if it is a result of differences in the test subjects and their levels of exposure to technology.

On the opposite end of the spectrum, however, is evidence that agents which are intrusive or annoying can generate an *affective distrust*, leading to disuse of the agent. As an example, Schaumburg (2000) performed a study of Microsoft Office users, showing that not only did subjects dislike the Office Assistant (or, as it was more commonly known, "Clippy"), but that they actually expressed strong negative feelings towards it. As a result, Clippy was ranked as the least efficient way to solve a problem, rejected in the context of learning a new application or feature (fewer than 33% said they would do so, 46% said they would never use him), and was only "liked" by 22% of subjects. Most subjects reported that they did not *trust* Clippy to correctly identify their goal or to provide useful assistance.

Trust vs. Risk

One additional point raised by Riegelsberger, Sasse and McCarthy (2005) in setting out their framework is that some researchers have shown that trust is only required in situations in which there is risk, although they claim that risk is hard to define. Generally, risk is measured economically, as the product of probability of success and gain (or, in cases where losses are likely, inverted cost) (Demaree, DeDonno, Burns & Everhart, 2008); however, this definition of risk is best applied to systems which are deterministic in nature (such as simple gambling tasks). Attempting to apply it as a metric in a trust framework results in a circular definition, in that it is the trust in the system which allows for the estimation of the probability of success. Social psychological measures of risk make use of game theory, resulting in a similar circularity. It does seem to follow, however, that risk is a function of the potential gains and potential losses of a given action or system, regardless of the actual form of that function. Thus, Riegelsberger et al.'s (2005) binary view of "risk" or "no risk" can be extended, meaning that the degree of trustworthiness required in any given interaction is proportional to the amount of risk the trustor must undertake.

In cases of distributed cognition, the trustor is not only making herself vulnerable (and thus, at risk) by not performing the entire task herself and with her own resources, and thus risking the outcome of this task, but she is also potentially wasting valuable cognitive resources and time as she learns to integrate the potentially untrustworthy artefact into her cognitive system.

Bridging the Gap

The current research about artefact use and coupling is highly contradictory. On the one hand, people appear to rapidly and automatically couple with artefacts (e.g. Kirsh, 2006; Kirsh & Magilo, 1994; Magilo & Kirsh, 1996; Clark, 1998; 2001a; Sutton 2006), and even to generate epistemic artefacts without being aware of doing so. On the other

hand, artefacts appear to be often misused or disused, even when the artefact is known to be more accurate (Lee & Moray, 1992; 1994; Dzindolet, Pierce, Beck & Dawe, 2002; Honeybourne, Sutton & Ward, 2006). The one point of agreement appears to be that some form of trust is required in order to create a coupled system; however, as I have shown previously, the ease with which that trust can occur is debated. One important distinction appears to have been missed in these debates, however: the difference in the very nature of the artefacts to which the coupling occurs.

Passive Artefacts

Passive cognitive artefacts, such as epistemic artefacts, are ancient and have evolved over time with humanity (Clark, 2001b), in a sort of evolutionary bootstrapping; tools allowed our forebears to be smarter, which allowed them to make better tools, in the same manner suggested by Chandrasekharan & Stewart (2007). This co-evolutionary process has, naturally, also left its mark on us; specifically, the availability of tools in our environment to perform the hard tasks necessary for success has made adapting to their discovery and use a better evolutionary strategy than attempting to overcome problems with our own limited resources. Passive artefacts are the tools that Dennett (1996) and Clark (2001a; 2001b) describe when they speak of offloading into the environment, both to free up cognitive resources and to allow us to reshape problems.

Thus, when dealing with such passive artefacts, Sterelny's hypothesized expensive social guards (Sterelny 2004, 2005) do not manifest themselves. It is unclear, however, if this is a result of Parsell's (2006) claims about the triviality of the cost of cheater detection or, if Sterelny was, in fact, correct about the high costs of employing social guards. From an evolutionary standpoint, the benefit of acquired behaviours needs to outweigh their costs, and so, it may the case that the expensive social guards were too complex to have evolved. It may simply be that the benefits of automatically endorsing the content of our epistemic artefacts far outweighed the costs of being deceived in a non-obvious and thus non-trivially detected way.

As a result of this automatic endorsement, the reliability criteria for coupling are met almost trivially, and thus people exhibit the sort of behaviour described by Sutton (2006) (and Clark, 1997, 2001a; 2001b; Kirsh & Magilo, 1994; et cetera), easily extending themselves to passive cognitive artefacts. The ubiquitousness of such artefacts in modern culture (notebooks, address books, paper, filing cabinets, palm pilots, et cetera) lends credence to this view.

Active Artefacts

Much more recently, however, there has been the creation of active cognitive tools: automated and semi-autonomous systems which are capable of manipulating representations. These are the systems which perform analyses and inferences, that make suggestions, that automate activities, and so on. As per Miller (2004), active tools have crossed the "agentification barrier,", and therefore, we treat them like agents, ascribing motivations, awareness and intent to these artefacts. These agent-like artefacts are sufficiently different from passive artefacts in that they do not induce automatic endorsement. Thus, Clark and Chalmers' (1998) concerns about the difficulty of meeting the reliability conditions in agent-agent interactions manifest themselves when interacting with active artefacts. As a result, we simply cannot create a coupled system with such an artefact until it has earned our trust. However, this process is made difficult by the fact that these artefacts lack many of the factors of trust which are employed in agent-agent interactions, such as temporal and social embeddedness and are markedly dissimilar from ourselves. And, of course, the amount of trust required in any given interaction or transaction is a function of the amount of risk undertaken by the trustor.

Thus, it is the offered reward (or, more accurately, the risk of getting less than the full reward), which explains the difference in results between the experiment Dzindolet, Pierce, Beck & Dawe (2002), in which people used their own judgement over that of an artefact they knew made fewer errors, and that of Gray et al. (Gray & Fu, 2004; Gray, Sims, Fu & Schoelles, 2006) in which reaction time was the only factor in the decision to use the artefact or not. In the latter case, the overall level of task performance was unimportant to the participants, and thus there was almost no risk in employing the artefact.

Multi-Function and Hybrid Artefacts

It is important to note that in the modern technological age, increasingly when discussing an artefact, we refer not to the physical device itself, but rather its software. For example, to a practiced user there is almost no functional difference between a physical or electronic address book. As such, both can be considered to be passive artefacts. The same can be said of many other software packages: electronic notepads, rolodexes and the like are all clearly passive devices. However, the same physical hardware which serves as an address book can also employ "active" software.

Some software artefacts, however, such as word processors, have begun to bridge the gap and act both as active and passive. Whereas older versions of such software simply allowed for the suspension of thoughts in linguistic form thus freeing us from our working memory limitations, newer ones alter the text we type by automatically correcting spelling and grammar, for instance. In general, I would suggest that such artefacts are true hybrids, and treated as such – being automatically trusted in their ability to hold our thoughts without being subject to alteration or error, while at the same time needing to earn our trust to be able to alter (or correct) them.

References

Camazine, S., Deneubourg, J.-L., Franks, N., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-Organization*

in Biological Systems. Princeton: Princeton University Press.

Chandrasekharan, S. & Stewart, T. (2007). The origin of epistemic structures and proto-representations. *Adaptive Behaviour, 15*, 329-353..

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

Clark, A. (2001a). *Mindware: An Introduction to the Philosophy of Cognitive Science.* New York: Oxford University Press.

Clark, A. (2001b). Reasons, robots and the extended mind. *Mind & Language, 16*, 121-145.

Clark, A. (2005). Intrinsic content, active memory and the extended mind. *Analysis, 65*, 1-11.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis, 58*, 7-19.

Demaree, H., DeDonno, M., Burns, J., & Everhart, D. E. (2008). You bet: How personality differences affect risk-taking preferences. *Personality and Individual Differences, 44*, 1484-1494.

Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness.* New York: Basic Books.

Dzindolet, M., Pierce, L., Beck, H. & Dawe, L. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44*, 79-94.

Dzindolet, M., Peterson, S. Pomranky, R, Pierce, L. & Beck, H. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies, 58*, 697-718.

Gray, W. & Fu, W-T. (2004). Soft constraints on interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*, 359-382.

Gray, W., Sims, C., Fu, W-T., & Schoelles, M. (2006) The soft constraints hypothesis: A rational approach to resource allocation for interactive behaviour. *Psychological Review, 113*, 461-482.

Hawkins, J. (2004). *On Intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines.* New York: Times Books.

Honeybourne, C., Sutton, S. & Ward, L. (2006). Knowledge in the Palm of your hands: PDAs in the clinical setting. *Health Information and Libraries Journal, 23,* 51-59.

Karowski, W. (2000). Symvatology: the science of an artifact-human compatibility. *Theoretical Issues in Ergonomics Science, 1*, 76-91.

Kirsh, D. (2006). Distributed cognition: A methodological note. *Pragmatics & Cognition, 14*, 249-262.

Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*, 513-549.

Lee, J. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*, 1243-1270.

Lee, J. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Maglio, P. & Kirsh, D. (1996). Epistemic action increases with skill. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Miller, C. (2004). Human-Computer etiquette: Managing expectations with intentional agents. *Communications of the ACM, 47*, 31-34.

Parsell, M. (2006). The cognitive cost of extending an evolutionary mind into the environment. *Cognitive Processing, 7*, 3-10.

Reeves, B. & Nass. C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.

Riegelsberger, J., Sasse, M. & McCarthy, J. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies, 62*, 381-422.

Schaumburg, H. (2001). Computers as tools or as social actors? – The users' perspective on anthropomorphic agents. *International Journal of Cooperative Information Systems, 10*, 217-234.

Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. In R. Schantz (ed), *The Externalist Challenge: New Studies on Cognition and Intentionality*. New York: Walter de Gruyter.

Sterelny, K. (2005). Made by each other: Organisms and their environment. *Biology and Philosophy, 20*, 21-36.

Sutton, J. (2006). Distributed cognition: Domains and dimensions. *Pragmatics & Cognition, 14*, 235-247.