

# Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing

BY EZEQUIEL A. DI PAOLO<sup>1\*</sup>, MARIEKE ROHDE<sup>1,2</sup> AND HIROYUKI IIZUKA<sup>1,3</sup>

<sup>1</sup>*Centre for Computational Neuroscience and Robotics, Department of Informatics, University of Sussex, Brighton, BN1 9QH, UK*

<sup>2</sup>*COSTECH, Université de Technologie de Compiègne, France*

<sup>3</sup>*Department of Media Architecture, Future University-Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido, 041-8655, Japan.*

{ezequiel, m.rohde, hi26}@sussex.ac.uk

\* *Corresponding author.*

**Abstract.** We introduce a series of evolutionary robotics simulations that address the behaviour of individuals in socially contingent interactions. The models are based on a recent study by Auvray, Lenay and Stewart (2006) on tactile perceptual crossing in a minimal virtual environment. In accordance, both the empirical experiments and our simulations point out the essential character of global embodied interaction dynamics for the sensitivity to contingency to arise. Rather than being individually perceived by any of the interactors, sensitivity to contingency arises from processes of circular causality that characterise the collective dynamics. Such global dynamical aspects are frequently neglected when studying social cognition. Furthermore, our synthetic studies point out interesting aspects of the task that are not immediately obvious in the empirical data. They, in addition, generate new hypotheses for further experiments. We conclude by promoting a minimal but tractable, dynamic and embodied account to social interaction, combining synthetic and empirical findings as well as concrete predictions regarding sensorimotor strategies, the role of time-delays and robustness to perturbations in interactive dynamics.

**Keywords:** Perceptual crossing, social contingency, agency, interaction dynamics, evolutionary robotics, simulation modelling. **PsycINFO Classification:** 3020, 3040

## 1. Introduction

In recent years, a dynamical systems turn has become increasingly popular in psychology and cognitive science (Beer, 2000; Port & van Gelder, 1995; Thelen & Smith, 1996). Some dynamical systems approaches attempt to capture observed psychological phenomena or theoretical constructs in terms of the properties of phenomenological mathematical models that describe a cognitive system in more or less qualitative terms (Kelso, 1995; Thelen & Smith, 1996; Van Geert, 1991). Others attempt to model minimal embodied systems from the ground-up; such generative models are not necessarily data-driven but cash out their scientific value in terms of the study of dynamical patterns observed and by linking these patterns to existing or new theoretical ideas (Beer, 1996, 2000; Webb, 1995). These two poles, the descriptive and the generative, define a continuum of dynamical approaches all of which go beyond the previous use of dynamical metaphors in psychology, (e.g., Heider, 1958; Lewin, 1951; Newcomb et al., 1952).

It is now widely acknowledged that investigating psychological phenomena in the context of situated interaction with an environment makes it possible to explain aspects of behaviour that are hard to grasp otherwise. This is especially true for the case of *social interaction* where two or more individuals are mutually coupled in perception–action loops. Their interaction can dynamically create phenomena that do not directly result from the individual capacities or behaviours of any of the partners if investigated on their own. However, to this date, most dynamical approaches to problems in social interaction have been located toward the qualitative end of the spectrum. Such models are sometimes disembodied in the sense that material and dynamical aspects of interaction are not taken into account. This has produced a series of models that concentrate on global patterns of interaction, (Granic, 2000; Vallacher et al., 2005; Van Geert, 1991, 2005) and consequently tend to be based on relational and global properties of the interaction process. There has been, in contrast, much less work in generative and embodied modelling of social interaction (Di Paolo, 2000; Iizuka & Ikegami, 2004; Quinn, 2001). The classical dichotomy between the priority of the social or individual levels of explanation plays a very practical role in such generative and embodied models, which leads to a fundamental theoretical question: If such models proceed first by constructing minimal individual systems which are only later put to interact socially, are these models not already committed to some form of methodological individualism? One of our aims in this paper is to explore this issue and answer this question negatively. We find that such models are well-suited to expose deeply rooted individualistic assumptions in social cognition precisely because they do not start by modelling dynamical patterns already constituted at the social level. As such, these models may become essential tools to take up the challenge of investigating the interaction process as a whole, rather than an individual’s capacities to behave socially in a social situation, and the dialectics between the individual and social levels of description, (De Jaegher, 2006).

In this paper, we explore the question of what makes an embodied social interaction distinct from other kinds of sensorimotor engagements with the environment. Is there anything about the dynamics of the interaction process that generates a specific temporal pattern that social interactors recognize and keep track of? Does the continuity of a social interaction depend on an individual’s capacity to recog-

nize the mutuality and contingency of the coupling? Or are there global dynamical structures of the whole social process that are sufficient for keeping an interaction under way?

Empirical evidence, such as Murray and Trevarthen's double TV monitor experiments and its successors (Murray & Trevarthen, 1985; Trevarthen, 1993; Nadel et al., 1999), indicates that individuals are not infinitely malleable and adaptable to the demands of an interaction if their partners do not themselves behave in a responsive manner. Two-month-old infants are able to interact with their mothers via a live double video link. However, when shown videos of their mothers generated during a previous interaction they do not engage in coordinations with the unresponding recording (which maintains intact the mother's expressiveness) and become distressed and removed. This seems to indicate that the recognition by the infant of the ongoingness and contingency of the interaction plays a fundamental role in its unfolding. Early involvement in socially contingent interactions, and its implied connectedness, plays a fundamental role in the infant's affective and experiential development (Tronick, 2004). Sensitivity to social contingency in two-month-olds is inferred from these results (Nadel et al., 1999). But this phrasing seems to suggest that such a 'recognition' is necessarily performed at the individual level. And this could be explained by individual-centred theories such as the postulation of an innate contingency detection module (Gergely & Watson, 1996). Our objective is to question this implication for the general case by showing how it is possible for the same results to stem from the rich dynamics of the social process itself. In this way, a dynamical analysis of minimal social interaction would bring support to social interactionist perspectives where the dyadic process is seen in itself as formative of individual social capabilities, (e.g., De Jaegher, 2006; Fogel, 1993; Stern, 2002). A recent empirical study brings some support to the social interactionist view (Markova & Legerstee, 2006), but the mechanisms involved in socially contingent affect attunement remain unclear. Our dynamical models intend to uncover some aspects of such possible mechanisms in conditions of minimal complexity.

We present a series of dynamical simulation models inspired by a recent study by Auvray, Lenay and Stewart (2006) on the dynamics of human perceptual crossing in a minimal shared virtual environment. This experiment can be seen as a variant of Trevarthen's double monitor experiments but simplified to the extreme. The details of their experiments are described in section 2. This is followed by a short presentation of the evolutionary robotics methodology used in this work and a discussion of its benefits and implications for psychology (section 3). In section 4, we introduce an evolutionary robotics model of the experimental paradigm which leads to a number of unintuitive insights about the data obtained. A second evolutionary robotics study (section 5) simulates the task in an extended set-up which again leads to the derivation of new hypotheses. In the conclusion (section 6) we summarise what the results from our models imply for the debate about dynamical systems approaches in psychology. We argue that a minimalist methodology, combining empirical psychological experiments and computer simulations, is a fruitful and promising methodological framework for investigating social behaviour as an embodied interaction process.

## 2. Perceptual crossing through tactile feedback

Auvray et al. (2006) (also Lenay, personal communication, 2006) have investigated the dynamics of human perceptual crossing in a minimal shared virtual environment. Two adult subjects, acting under the same conditions, can move a cursor left and right along a shared one-dimensional virtual tape that wraps around. They are asked to indicate the presence of the other partner. The subjects are blindfolded and all they can sense are on/off tactile stimulations on a finger when their cursor crosses an entity on the tape. Apart from each other, participants can encounter a static object on the tape, or a displaced ‘shadow image’ of the partner, which is strictly identical to the partner as regards to size and movement characteristics (see Fig. 1). All objects are small and of the same fixed size. Subjects are told that they can encounter a moving object, a static object, and another sensing subject and are asked to click the mouse button when they think they are scanning the other subject.

The problem is, therefore, not only to distinguish moving and static entities along the tape, but to distinguish two entities that move exactly the same way, only one of which represents the other subject, who can sense the first subject’s presence and respond contingently to her actions. In this sense, these experiments are, even though much simpler, analogous to Trevarthen’s experiment: The only difference between the mother’s behaviour on the monitor between the conditions is whether she senses the child and reacts to it or not; her expressive behaviour, i.e., her motion, language, mimics, dynamics, voice etc. are identical between the two conditions.

[Insert Figure 1]

The results by Auvray et al. show that subjects are, despite the poor sensory information provided by the minimalist virtual environment, very successful at solving the task ( $\approx 70\%$  correct responses) without the need of training. How is this discrimination achieved? Even though the behavioural patterns for long-term interaction with objects or subjects vary significantly across individuals and trials, a frequent strategy adopted initially by subjects is to make oscillatory scanning movements around any entity sensed on the tape.

A closer look at the data reveals that subjects are not able to make a conscious distinction between the shadow image and the interacting partner. Sometimes, if one of the partners scans the fixed object, the shadow image mirrors this scanning activity, and the other subject can interpret this scanning as oriented towards her. This is represented by the fact that the relation between caused stimulation and recognition mouse click response is approximately equal for both actual perceptual crossing and crossing of the shadow image. In other words, subjects are equally likely to click their mouse when scanning another subject or her shadow. For both cases, the amount of clicking is much higher than it is for the crossing of the fixed object, which seems to be hardly ever mistaken for the other subject.

So, if the movement of a subject and its shadow are indistinguishable, and the mouse click response occurs with similar probabilities when scanning another subject or scanning his shadow, how can we explain the 70% successful responses? According to Auvray et al., the answer lies in understanding the dynamical properties of the interaction process as a whole. The important issue is that the scanning

of an entity encountered will only *stabilise* in the case that both partners are in contact with each other – if interaction is only one-way, between a subject and the other’s shadow, the shadow will eventually move away, because the subject it is shadowing is still engaged in searching activity. Two-way mutual scanning is the only *globally stable condition*. Therefore, the solution to the task does not rely on individuals performing the right kind of perceptual discrimination between different momentary sensory patterns, but emerges from the mutual perceptual activity of the experimental subjects that is oriented towards each other.

Interesting hypotheses and further questions can be deduced from these results, for instance, whether such global dynamical processes, in which none of the individual actors can be held responsible for the interactional success, also play a role in more complex conditions, such as Trevarthen’s double monitor experiments. But also within this minimal experimental set-up, there are more questions to be asked. In this paper, we investigate simple simulated robotic agents performing the same task from a dynamical systems perspective. Due to the novelty of the approach we must first detour to explain how such purely synthetic findings can enrich the practice of the experimental psychologist.

### 3. Minimal cognition and evolutionary robotics

Evolutionary Robotics (ER) is a “technique for the automatic creation of autonomous robots [...] inspired by the darwinian principle of selective reproduction of the fittest” (Nolfi & Floreano, 2000, preface). The experimenter determines some aspects of the robot’s or simulated agent’s architecture, but underdetermines others. These are left to be determined in an automated way by an evolutionary search algorithm, according to the optimisation of an abstract performance measure (fitness function). For instance, if mobile agents must locate and approach a light source, fitness could be defined as the proximity to the light source at the end of each test trial. Initially, all parameters describing the agent’s controller have random values, typically resulting in low or zero scores for the generated behaviour. However, the agents that are slightly better than the rest will be selected to generate offspring, similar to themselves, to feed the next generation of agents. For the case of light-seeking, for instance, those agents that randomly move a little bit in the direction of the light will receive a higher score than those that do not move at all. Their ‘offspring’ are again evaluated, and after a number (hundreds, sometimes up to thousands) of repetitions of this cycle, the population of agents will perform the desired behaviour, even though the human designer never explicitly programmed the solution (see figure 2).

[Insert Figure 2]

What is interesting about this method is that the performance criterion does not directly evaluate the control function of the agent, but its *embodied behaviour* within a closed sensorimotor loop. The automated search can exploit aspects of the complex dynamical interaction between the artificial agent and its simulated environment, and these solutions are typically the ones that human designers do not think of when explicitly programming an agent to perform a certain behaviour. For the case of light seeking behaviour, many solutions are possible and equally viable, e.g., using one light sensor or many, approaching the light forwards and backwards,

stopping once there or cycling around it, etc. The evolutionary search can come up with any of these solutions, while a human designer would only find few of them intuitive. Indeed, it happens frequently that the evolutionary roboticist finds it difficult to understand the evolved behavioural solutions. In these cases, a ‘pseudo-empirical’ investigation of the agents follows in order to explain their performance: Agents are tested under different psychophysical conditions, internal and external variables are monitored, the structure of the evolved agent control architecture is closely examined or altered, etc.

Typically, the systems that are designed in evolutionary robotics are controlled by continuous-time recurrent neural networks (CTRNNs, (Beer, 1990, 1995)). These neurocontrollers are particularly useful for dynamical models since they allow the specification of multiple timescales, from the very fast to the very slow, including behavioural, learning and developmental timescales in a single neural network (see appendix).

These models are not strict models in the traditional sense of trying to fit an empirically generated data set. They serve as tools for grounding and questioning preconceptions held about the explanandum, i.e. how a certain behaviour is generated. Most other methods in artificial intelligence modelling do not include this explicit self-critical factor. Typically, the models generated with this approach are deliberately minimal (Beer, 2003; Harvey et al., 2005). The manageable complexity of these models, together with the fact that, in general, all variables and parameters can be accessed and recorded, makes it possible to fully understand and explain the dynamics of agent behaviour. Some examples of this methodology have recently been applied to modeling animal cognition, (e.g., Vickerstaff & Di Paolo, 2005). Beer (2003) argues that the minimalism of this method allows us to perform the necessary mental gymnastics to deal with real, dynamical, and context-dependent cognitive performance.

But how can artificial systems that are deliberately minimal in complexity and therefore not even remotely comparable to human beings be of scientific use to the psychologist? As stated above, these systems should not be seen as models in the strict sense. They are not models of humans, but they can help to better understand a very particular and limited aspect of human sensorimotor behaviour. Experiments with human subjects that focus on particular aspects of behavioural dynamics, such as the study by Auvray, Lenay and Stewart described above are, therefore, very suitable for ER modelling. The empirically investigated variables, sensation and motion on the tape over time, can be fully captured in a simple simulation, without drawing any conclusions about the relation between structures that give rise to this behaviour in humans or simulated agents.

#### 4. Modelling perceptual crossing

We propose a model of the original experiment by Auvray et al. which provides a more thorough understanding of the task. We briefly outline the technical aspects of our simulation studies, see appendix for technical details. The original experimental settings are included in the model, i.e., agents can meet a shadow image of the partner on the tape, the partner itself or a fixed object (all of which have the same size on the tape).

The agents can move along one dimension, i.e., to the left and to the right. This one-dimensional world wraps around. The agents are controlled by CTRNNs. They have one touch sensor that feeds an on/off signal into the network if they touch the other agent or any other object located on the tape. The output activation of the neural network is used to control the left/right movement of the agent. The parameters and architecture of this recurrent network are evolved with a genetic algorithm to maximise the performance of the task which is to locate the partner agent and spend as much time as possible as close to each other as possible while not being trapped by static objects or shadow images. In this simulation, both partners are identical, i.e., just a single population of agents is evolved.

When we first tried to evolve agents to solve the perceptual crossing task, the evolutionary search algorithm was not able to find a satisfactory solution. The behaviour that evolved was for agents to halt when crossing any object encountered on the tape, be it the partner, the fixed object or the shadow of the other. Given the experimental set-up, this is a comparably successful strategy: If agents first encounter each other, or if one agent runs into its waiting partner, it achieves perfect fitness, and these are the majority of possible cases. However, it is neither the *optimal* behaviour, as in the remaining cases, the agents will not find each other at all, nor is it a very intelligent or adaptive solution and it does not resemble any of the strategies adopted by human subjects, who keep actively exploring. Only when a small time delay between a crossing on the tape and the agent's sensation was included into the model (see appendix), the evolutionary search algorithm came up with an adaptive solution. The trajectories generated by the agents are similar to those generated by some human subjects (Fig. 3 (C)).

An interesting question arises from these unsuccessful trials: Is the oscillating scanning movement that most subjects adopt initially related to the existent delays between sensation and reaction in humans? It seems natural to us that subjects would adopt a strategy such as oscillatory scanning, however, it is not *a priori* necessary, and it even seems like a waste of energy. There are many possible explanations for this behaviour. However, one hypothesis that we can derive from our model is that sensory delays may play a role in generating scanning behaviour, and that subjects, possibly, repeatedly cross the partner because of reaction time delays, like the evolved agents do. This hypothesis would predict that the amplitude of scanning oscillations around a target will be positively correlated with the amount of sensorimotor delay, which can be tested in further empirical experiments.

[Insert Figure 3]

Taking a closer look at the evolved solution to the task, an interesting observation is that the discrimination between the shadow image and the other subject evolves quickly whereas it seems more difficult to evolve agents that can distinguish another agent from a fixed object. These findings contradict the intuition that the easier task would be to recognize a static object, while distinguishing two moving objects, only one of which is in turn responsive, seems much harder, particularly because their motion characteristics are identical<sup>1</sup>. A close look at the data from the simulation model, however, changes the picture.

<sup>1</sup> A possible explanation of the apparent ease with which humans distinguish fixed objects would make use of the invariant correlation that it is possible to build between tactile and proprioceptive sensory input during active scanning. This possibility is also available to the agents

In the simulation, there is a striking similarity between sensorimotor patterns for coordinated mutual scanning (crossing) and for scanning a fixed object (Fig. 3 (A) and (B)). Encountering any stimulus makes the agent revert its direction of movement, which leads to another encounter followed by another inversion of velocity, and so forth. If both agents engage in this motion pattern, they will always meet in the middle of their return trajectory. This coordinated activity leads to sensations and motions changing over time in a way very similar to those that come about when investigating a fixed object. When we inspect the duration of the stimulus upon crossing a fixed object, we realise that it lasts longer than when crossing a moving partner. This is because the fixed object does not move itself. The solution that the simulated agent adopts simply relies on integrating sensory stimulation over a longer time period, which yields a higher value for a static object, i.e., it is sensed as having a larger apparent size. This hypothesis has been tested and is supported by the fact that the agent is quite easily tricked into making the wrong decision if the size of the static object is varied, i.e. a small object is mistaken for another agent — or, likewise, the other agent is perceived as a smaller object.

It is interesting to note that the smaller perceived size in the case of perceptual crossing depends on encounters remaining in anti-phase oscillation (figure 3 (A)), which is an *interactionally coordinated property* (De Jaegher, 2006). A systematic distinction between objects having the same objective size is therefore *co-constructed* by the agents during coordinated interaction. Objects systematically moving in opposing velocities to the scanning movements will appear smaller in size. The co-construction of this appearance resides in the systematic opposition of velocities which is a coordinated relational property of the interaction. In turn, individuals respond to this emergent coordination by remaining in coordination with the apparently smaller object.

These simulation experiments do not directly tell us anything about the strategy humans employ in this task. There are many viable solutions to the task, and it is rather unlikely that humans use the strategy just described, as it appears rather specific to the conditions under which the agents were evolved. So what do they imply? First of all, they demonstrate that an approach that does not just look at the individual capabilities, but also at phenomena that emerge during embodied and situated interaction uncovers relevant factors that are easily overlooked otherwise. A task that intuitively seems difficult, i.e. to distinguish two entities with identical movement characteristics (the partner and the shadow image), becomes almost trivial, if the effects emerging from the mutual search for each other are taken into consideration. This finding already results from the minimal empirical closed loop experiments by Auvray et al. (2006). On the other hand, the intuitively easy task of distinguishing a moving entity (the partner) from a static one is indeed non-trivial, if the emergent effects of interaction, i.e. anti-phase coordination, are taken into consideration.

Furthermore, this issue is uncovered by the evolutionary robotics simulations prompts a closer look at the empirical data which already confirms that subjects spend approximately twice as much time crossing a fixed object ( $\approx 33\%$  of stimulations) than they do crossing the other subject's shadow ( $\approx 15\%$  of stimula-

---

in the model since the neurocontroller allows for recurrent feedback to be used should evolution find that this is an advantageous strategy.

tions)<sup>2</sup>. It therefore seems that the distinction that arises from interaction (which moving object is the partner?) is more efficiently solved than the distinction that requires subjects to individually succeed in recognition (which object that stays in my neighborhood is the fixed object?). This empirical finding is compatible with the implications of the minimal model’s results.

With this global view on the dynamics of perceptual crossing in the investigated set-up, these insights may seem almost trivial. However, if we had started from the perspective of the individual and its conscious recognition capacities, these findings would be mysterious: Why would subjects spend relatively more time scanning the fixed object if it is so much easier to identify?

## 5. The dynamics of minimal agency detection

The findings of Auvray et al. point to the dynamics of the interaction process as the explanation of coordinated crossing between subjects and not to an individual sensitivity to social contingency. The model presented in the previous section confirms this result. What remains unclear is the generality and significance of this finding. Should we expect interactive dynamical factors to account for differences of behaviour between other contingent and non-contingent situations? Already in infants it seems proper to describe the reactions to a pre-recorded social stimulus as indicators of an *individual* capacity to discriminate it from live interactions. But the previous results suggest that the stability of the interaction dynamics may be a factor which is co-present and possibly constitutive of such an individual awareness of a non-contingent social situation. However, if – as in the above case of unimodal perceptual crossing – all that matters is the fact that coordinated interaction is the only long-term stable dynamical mode, then it seems that the situations where the social process itself would bear some responsibility for the difference between contingent and non-contingent behaviour would be rare. We can expect multi-modal, noisy social encounters to have several and time-varying modes of unfolding. In order to further explore this issue, we propose a variation on the above model that will help clarify the possible role of the interaction dynamics in producing a difference of behaviour between contingent and non-contingent situations by investigating a set-up in which the interaction of one agent with another’s nonresponding recording is unconstrained.

One important difference between the setup investigated by Auvray et al. and Murray and Trevarthen’s double TV monitor experiments is the simultaneous presence of contingent and non-contingent sources of stimulation (the other subject’s sensitive position and its shadow). In this section we present a model where the two situations are separated into distinct live and replayed interaction phases – a situation only partially explored in a previous evolutionary robotic model (Di Paolo, 2000). We investigate how embodied agents can establish a live interaction and discriminate this from an interaction with an unreactive agent performing pre-recorded movements from a previous interaction. The issue we want to test is *whether such a discrimination requires a complex perceptual strategy on the part of the discriminating agent or whether simpler solutions emerge from the interaction process itself*.

---

<sup>2</sup> Even though, as stated above, the crossing of the moving object is much more likely to trigger a recognition response from the subject.

We use almost the same settings as in the previous model except for the number of sensors which is now increased to six in order to allow for more accurate discrimination of the partner’s movements (this would facilitate individual discrimination strategies if they were to be favoured by the artificial evolutionary process). The static objects and shadow images are removed (the only perceivable object is the other partner) and the one-dimensional space is bounded to the left and right with walls. A schematic view of the set-up is shown in Fig. 4.

[Insert Figure 4]

As before, agents are controlled by a CTRNN, all parameters of which are artificially evolved. If both partners were evolved from a single population, we would be exploring the special case of an interaction between twinned partners. To avoid this, partners are co-evolved by using two populations of agents (one labeled the top agent and the other the bottom agent). The fitness is calculated on the basis of two factors. The first is how many times the agent can cross its central position with that of a live interacting partner (live interaction). The other is how much the agent can stay away from a dummy, non-contingent partner which only replays the motions of its partner as recorded in the live stage (one-way interaction). Agents retain no memory after either test. In this way, the motions of the dummy and the live agents are the same and the situation resembles that of the double TV monitor experiment. In our current settings, only the bottom agent is required to discriminate the live interaction from the one-way interaction. Therefore, the top agent is evaluated only with respect to the first factor (it always tries to engage in coordinated crossing). For the bottom agent, the fitness is calculated by averaging the first and second factors. (See appendix for technical details.)

Evolved agents successfully discriminate between the two conditions. Figure 5 shows the movement trajectories for both agents (live interaction) and the movement of the bottom agent when starting from a different initial position and interacting with the recording of the top agent (one-way interaction). At the start of the interaction, the bottom agent behaves similarly in both cases. During this phase, the positions of the agents cross rarely and both move towards one of the walls. Afterwards, in the live interaction, the crossing coordination is established by the bottom agent moving away from the wall a little, while the top agent oscillates around the position of the bottom agent. The top agent ‘catches’ this movement of the bottom agent and provokes a return to the wall through its oscillatory interaction. By contrast, in the one-way interaction case, a slight difference of the timing of the movement away from the wall is enough for the recorded trajectory of the top agent not to be able to ‘respond’ in time to ‘catch’ the bottom one. This event causes the breakdown of the coordination and the lack of mutuality in the interaction makes this breakdown irrecoverable.

It is also found that the same divergence of behaviour happens even if the initial positions (and neural states) of the bottom agent are exactly the same in the live and one-way interaction, since slight differences in trajectories due to sensor noise accumulate. The performance of the agents is robust to noise during live interaction and they can sustain coordinated behaviours, while this is not the case for a single agent in the one-way interaction and this causes the coordination to break down eventually. The evolutionary algorithm has found a solution where the mutual coupling results in a feedback loop where coordination (crossing) is robust

to perturbations. But, the feedback is broken in the one-way interaction and so the lack of robustness itself causes a behavioural change that ends up in the bottom agent moving away from interaction. To support this hypothesis figure 6 shows the stability of the coordinated behaviours in the case of both live and one-way interactions as a function of the level of sensor noise. It is clear that the noise can be filtered out when both agents interact – resulting in sustained coordination – but its effect on movement is amplified and breaks the coordination in the one-way interaction.

[Insert Figure 5]

[Insert Figure 6]

This simple model answers our question. It is sufficient to exploit dynamical properties of the mutuality of live interaction, such as increased robustness to noise and timing differences, to produce a different behaviour in the case of a one-way interaction. The results suggests that the recognition of social contingency needs not lie in complex cognitive individual modules able to integrate past information, but rather in the situated ongoingness of the interaction process itself, in its dynamic landscape, and its robustness to perturbations.

## 6. Discussion

Social contingency is a relational condition, and its perception requires the evaluation of not just the patterns of expressive action of a partner, but also of how these relate to our own actions. How would it otherwise be possible for infants to distinguish a live interaction with their mothers from a recording of a previous one? The perception of social contingency is therefore the appreciation of an inherently interactive property of a dynamical process of reciprocal causality. This fact has not stopped the proposal of descriptions and explanations of such perceptual capacities in terms that are more individualistic than properly interactive. Infants in the double TV monitor experiments are described as themselves being sensitive to social contingency (Nadel et al., 1999), as if the right comparisons, associations and inferences could only be made by innate computational modules in the infant’s head. The problem of this situation is that such individualistic explanations are currently not properly balanced by an exploration of alternative hypotheses that put the dynamics of the interactive process in a more central role. Despite empirical support for social interactionist alternatives (Markova & Legerstee, 2006) properly social mechanisms are not often put forward as explanations of contingency. One of the reasons for this may partly be because the dynamical systems tools that are necessary to explore the space of rich, multi-timescale, polyrhythmic, multi-modal interpersonal interactions are still under development. Hence our focus on experiments and models that minimize complexity. Not because they will necessarily uncover all the relevant theoretical insights required to understand the more complex natural situation, but because they open up the space of explanations available and (in the spirit of Beer’s (2003) minimal cognition) they demand the engagement with the right kind of ‘mental gymnastics’.

In order to go beyond paying lip-service to the relevance of interaction and feedback, our models show in concrete terms what aspects of the interaction dynamics

are responsible for the response to social contingency. The perceptual crossing model supports the explanation proposed by Auvray et al. Participants in the empirical study were unable to distinguish between responsive partners and their shadows in terms of the pattern of sensory inputs. And yet, since shadows do not remain on the spot for long, the only global stable mode is when partners are crossing each other, a discrimination that arises from interaction with similar ease in our simulation model. But the results also uncover non-trivial possibilities regarding the discrimination between live partners and fixed objects in terms of co-constructed modulation of perceived size. This, again, demonstrates, how a socially coordinated dynamic pattern results in a change to individual perception. In particular, we hypothesize that, in the human experimental data, whenever a subject is making a decision to remain in interaction with an object or discard it as a static one, one of the predictors for this decision will be the total time of previous stimulation (i.e., the integrated apparent size of the object). A global view of the interaction process allows us to formulate this prediction since such a view provides access to phenomena and variables that are too easily overlooked otherwise.

The second model reproduces the experimental setting of live vs. recorded interactions more closely whilst maintaining minimalism. Consequently, it enables the exploration of other social dynamical explanatory factors: The stability of the mutual feedback against perturbations of timing and noise, and the active co-construction of an interaction process that is sensitive to social contingency. The self-correcting aspects of the interactive process simply break down for an agent that interacts with a recording. Under these circumstances, it is simply unnecessary to evolve an individual contingency recognition strategy, the social process takes care by itself of inducing the individuals to produce the right behaviour.

In both our models, discrimination between contingent and non-contingent situations is achieved through the social process in the ongoingness of the interaction. The dynamics of interaction are not simply the data that an individual must evaluate, but are the evaluation process itself. There is no *a priori* reason, at the moment, to assume that explanatory possibilities have to be either strictly individual based or strictly socially modulated. Possibly, many appropriate explanations for socially interactive processes incorporate both factors and, thereby, lie somewhere along a spectrum, defined by strictly individual evaluation of interactive information at one end and purely social modulation of individual dynamics on the other. What is called for is a methodology that would allow to map this spectrum by 1) determining the dynamical properties of coordination present in a given social interaction and 2) generating hypotheses regarding their contribution to the observed social behaviour. Such tools would also allow the exploration of the mutual shaping (as well as the tension) between individual and social dynamics as an intrinsic source of de-stabilization of coordination. A concrete prediction generated by this work is that interactive factors affecting coordination may be uncovered by their signature response to controlled perturbation methods. The expected robustness to noise for those coordinations that stem from social dynamics would be qualitatively closer to that indicated by figure 6.

The combination of empirical studies and simulation models illustrates how both methods can inform each other: A psychological study is modelled in a computer simulation, which increases our understanding of the data obtained thanks to the less prejudiced design afforded by the evolutionary robotics methodology. From

these results, an extended version of the experiments is generated, which is first investigated in simulation, leading to refined hypotheses and ideas. These ideas will subsequently be tested in empirical psychological experiments. The typical complaint when confronted with artificial models of the dynamics of psychological process is that of the gross gap in complexity between the model and the modelled situation. This problem is solved in this work by pairing up an empirical experiment and a computer model that both deliberately strive for minimalism, in the spirit of keeping the dynamics of the investigated behaviour tractable. The fact that even such minimal models lead to unintuitive findings speaks for the dynamical complexity of the subject matter and the usefulness of the approach. We argue that such two way interaction between minimal dynamical simulation models and minimal dynamical psychological experiments is likely to be fruitful for a larger class of scientific problems.

**Acknowledgments:** This research was partially supported by the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for JSPS Fellows, 17-04443.

## Appendix: Technical details

### (a) Continuous-time recurrent neural network

A neuron  $i$  in a CTRNN is governed by the following continuous-time differential equation:

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^N w_{ji} z_j(y_j) + I_i, \quad z_i(x) = 1/(1 + e^{-x-b_i}),$$

where  $y_i$  represents the cell potential,  $z_i$  is the firing rate,  $\tau_i$  is its time constant (modulating the speed of response of the node),  $b_i$  is a bias term, and  $w_{ji}$  is the strength of the connection from the neuron,  $j$ , to  $i$ .  $I_i$  represents the sensory input, which is given to only sensory neurons. The number of neurons is given by  $N$ .

### (b) Details of the model in Section 4

A generational genetic algorithm with truncation selection ( $\frac{1}{3}$ ) and a real valued  $\in [0, 1]$  genotype with vector mutation (Beer, 1996) is used. The agents and all other objects have a size of four units, the tape size is 600 units. The CTRNN controllers that has four input neurons, corresponding to four neighbouring binary touch sensors, up to 5 hidden neurons and two motor neurons for left and right movement. The motor neurons are treated as hidden neurons apart from the fact that they generate the motor values, i.e., input neurons can connect to them directly and they can form recurrent connections with themselves or other hidden neurons. The parameter ranges are  $b_i \in [-3, 3]$ ,  $\tau \in [20, 3000]$  (in timesteps), and  $w_{ji} \in [-8, 8]$ .  $\tau$  is mapped exponentially from the genome value  $\in [0, 1]$  to the target range, the other parameters were mapped linearly. The network structure is partially evolved (step functions  $x > 0.7$  and  $x > 0.6$  decide about the existence of connections and hidden neurons respectively). A sensory gain  $S_G$  and a motor gain  $M_G$  where also

evolved, both of which were mapped exponentially to the range  $[1, 100]$ . Agents are tested against clones of themselves using a weighted fitness average over six trials. The fitness criterion is the average distance from the other

$$F = \frac{1}{T} \sum_0^T 1 - \frac{d(t)}{300}.$$

The sensory delay used was 50 time steps. The trials lasted  $T \in [8000, 11000]$  time steps.

(c) *Details of the model in Section 5*

The CTRNN used here consists of 10 neurons, including 6 input and 2 output neurons. All neurons are fully connected. The range of all parameters are set to  $b_i \in [-3, 3]$ ,  $\tau \in [1, 100]$ , and  $w_{ji} \in [-8, 8]$ , respectively. These parameters are evolved using a rank-based genetic algorithm with elitism. In order that the top and bottom agents can acquire different behaviour strategies, we used a co-evolutionary system which has 2 populations, which are assigned to top and bottom agents. Each population consist of 20 agents, which are evaluated on the basis of the results of all  $20 \times 20$  combinations between populations. Only vector mutation operators (Beer, 1996), which adds a small random vector to the real-valued genotype, are used. The best 6 agents of the population are retained and the others are replaced by the agents with mutation which are selected from the original populations based on a fitness. The different fitness function,  $F^1$  or  $F^2$ , is given to each population as follows.

$$F_i^1 = \frac{1}{20} \sum_{j=1}^{20} \#Cross(a_i^1, a_j^2) \times (20 - \#Cross(a_i^1, Record(a_j^2)))$$

$$F_i^2 = \frac{1}{20} \sum_{j=1}^{20} \#Cross(a_j^1, a_i^2)$$

where  $\#Cross(a_i^1, a_j^2)$  represents the number of crossing in the interactions between agents,  $a_i^1$  and  $a_j^2$ , but the maximum number is set to 20 in order to avoid unreasonable oscillation.  $Record(a_j^2)$  means the recorded motions of agent,  $a_j^2$ . Therefore, the more agent  $a_i^1$  crosses with the recorded motions, the less the fitness become.

## References

- Auvray, M., Lenay, C., & Stewart, J. (2006). The attribution of intentionality in a simulated environment: the case of minimalist devices. In *Tenth Meeting of the Association for the Scientific Study of Consciousness, Oxford, UK, 23-26 June, 2006*.
- Beer, R. D. (1990). *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroscience*. San Diego: Academic Press.
- Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3, 469-509.
- Beer, R. D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In Maes, P., Mataric, M. J., Meyer, J.-A., Pollack, J. B., &

- Wilson, S. W. (Eds.), *From Animals to Animats 4: Proceedings of the 4th International Conference on Simulation of Adaptive Behavior*, pp. 421 – 429. MIT Press, Cambridge, Mass.
- Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science*, 4, 91–99.
- Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11, 209–243.
- De Jaegher, H. (2006). Social interaction rhythm and participatory sense-making. An embodied, interactional approach to social understanding with some implications for autism. Unpublished DPhil Thesis, University of Sussex.
- Di Paolo, E. A. (2000). Behavioral coordination, structural congruence and entrainment in acoustically coupled agents. *Adaptive Behavior*, 8, 27 – 47.
- Fogel, A. (1993). *Developing through relationships: Origins of communication, self, and culture*. Chicago: University of Chicago Press.
- Gergely, G., & Watson, J. L. (1996). The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control in infancy. *International Journal of Psycho-Analysis*, 77, 1181–1212.
- Granic, I. (2000). The self-organization of parent-child relations: Beyond bidirectional models. In Lewis, M., & Granic, I. (Eds.), *Emotion, development, and self-organization. Dynamic systems approaches to emotional development*. Cambridge: Cambridge University Press.
- Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M., & Tuci, E. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11, 79–98.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.
- Iizuka, H., & Ikegami, T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, 10, 361–378.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge MA: MIT Press.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper and Brothers.
- Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology*, 42, 132–141.
- Murray, L., & Trevarthen, C. (1985). Emotional regulations of interactions between two-month-olds and their mothers. In Fox, T. M. F. . N. A. (Ed.), *Social perception in infants*, pp. 177–197. Norwood, NJ: Ablex.
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, 2, 164–174.
- Newcomb, T. M., Turner, R. H., & Converse, P. E. (1952). *Social Psychology: The study of human interaction*. London: Routledge.
- Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics. The biology, intelligence, and technology of self-organizing machines*. Cambridge MA: MIT Press.
- Port, R., & van Gelder, T. (1995). *Mind as motion: explorations in the dynamics of cognition*. Cambridge MA: MIT Press.
- Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J., & Sosik, P. (Eds.), *Advances in Artificial Life: Sixth European Conference on Artificial Life*, pp. 357–366. Berlin: Springer Verlag.

- Stern, D. (2002). *The first relationship: Infant and mother*. Harvard University Press.
- Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Trevarthen, C. (1993). The self born in intersubjectivity: The psychology of an infant communicating. In Neisser, U. (Ed.), *The Perceived Self*, pp. 121–173. Cambridge: Cambridge University Press.
- Tronick, E. Z. (2004). Why is connection with others so critical? The formation of dyadic states of consciousness: Coherence governed selection and the co-creation of meaning out of messy meaning making. In Muir, J. N. . D. (Ed.), *Emotional development*, pp. 293–315. Oxford: Oxford University Press.
- Vallacher, R. R., Nowak, A., & Zochowski, M. (2005). Dynamics of social coordination: The synchronization of internal states in close relationships. *Interaction Studies*, 6, 35–52.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98, 35–52.
- Van Geert, P. (2005). Vygotskian dynamics of development. *Human Development*, 37, 346–365.
- Vickerstaff, R., & Di Paolo, E. (2005). Evolving neural models of path integration. *Journal of Experimental Biology*, 208, 3349–3366.
- Webb, B. (1995). Using robots to model animals: a cricket test. *Robotics and Autonomous Systems*, 16, 117–134.

## Figure captions

**Figure 1:** Schematic diagram of the one-dimensional environment in perceptual crossing experiment.

**Figure 2:** Illustration of the evolutionary cycle in evolutionary robotics.

**Figure 3:** Results from evolved model. (A) Stabilised perceptual crossing (trajectories and sensorimotor values). (B) Scanning of a fixed object (trajectories and sensorimotor values). (C) The trajectories are very similar to those generated by human subjects (lighter lines: shadows/static objects). All diagrams include motor noise.

**Figure 4:** Schematic view of the environmental settings. The agents can move horizontally and sense a body of the other agent with six arrayed tactile sensors.

**Figure 5:** Trajectories of agents over time. The dotted and thick solid lines show the trails of the top and bottom agents in the live interaction, respectively. The thin solid line shows the trails of the bottom agent when it interacts with the nonresponding recording which replays the dotted line.

**Figure 6:** Stability of coordinated behaviours against noise strength. The average number of crossing over 100 trials with random initial positions is shown. The solid line and the dotted line represent the cases of the live and the one-way interaction, respectively.

### Figures

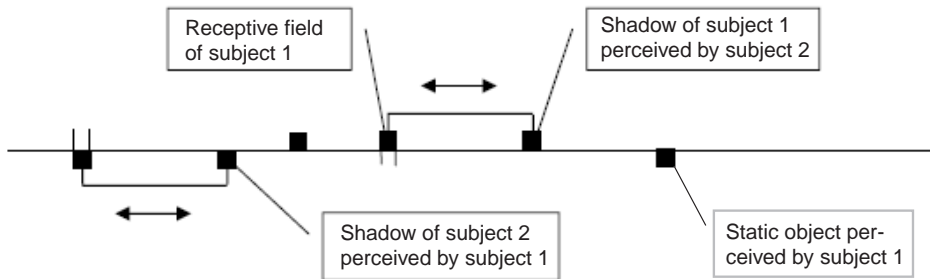


Figure 1.

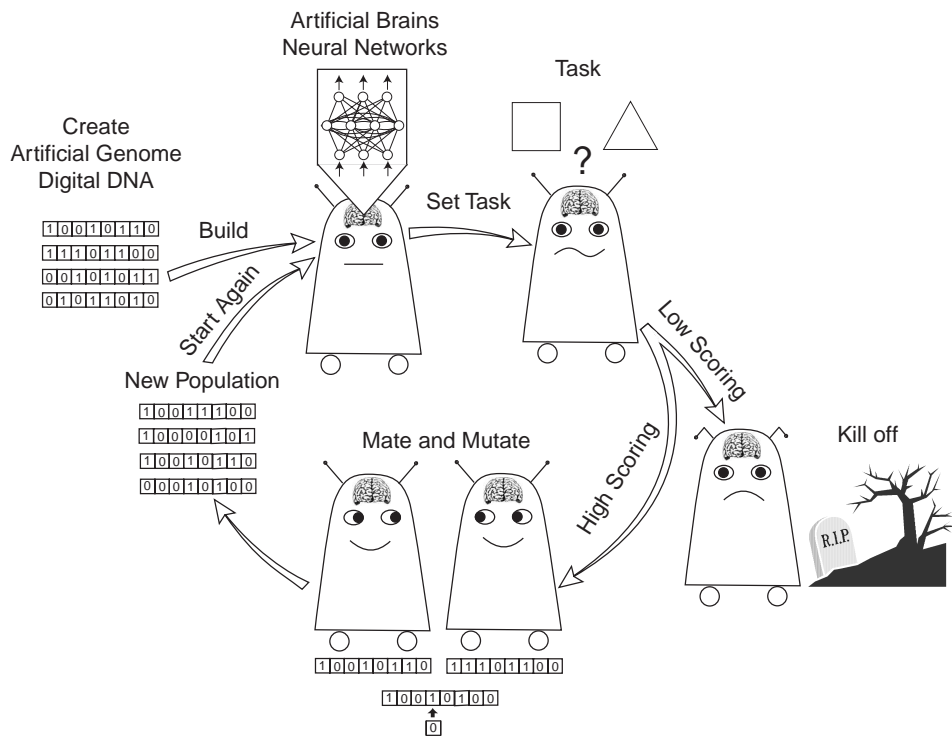


Figure 2.

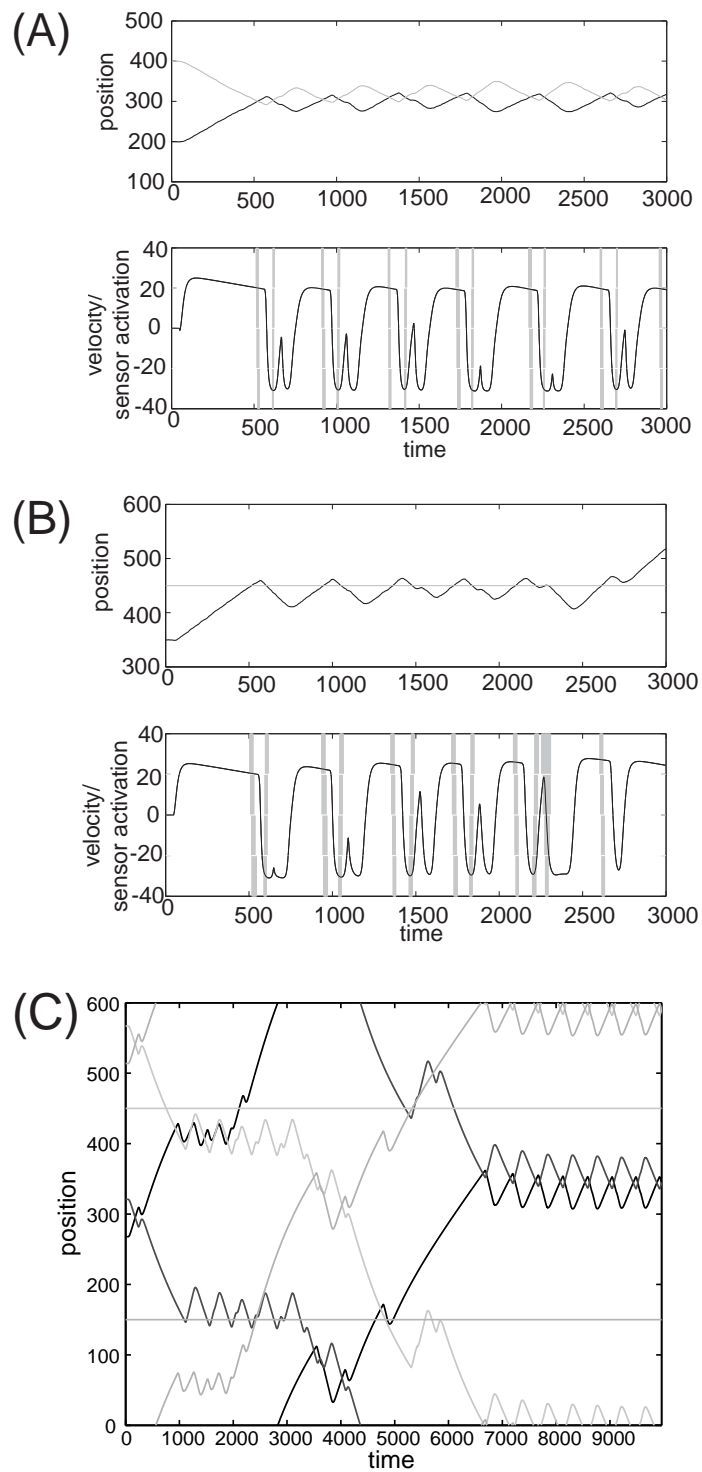


Figure 3.

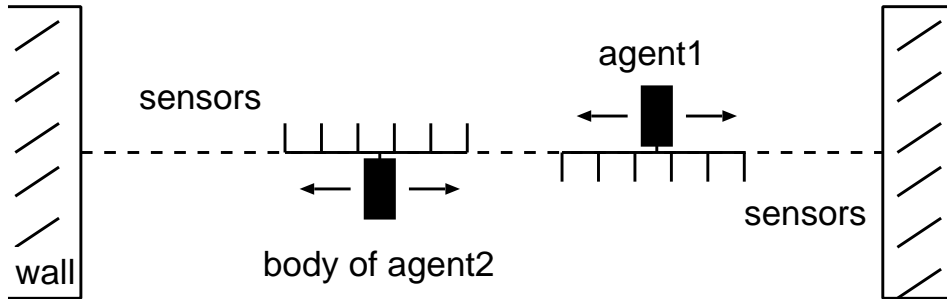


Figure 4.

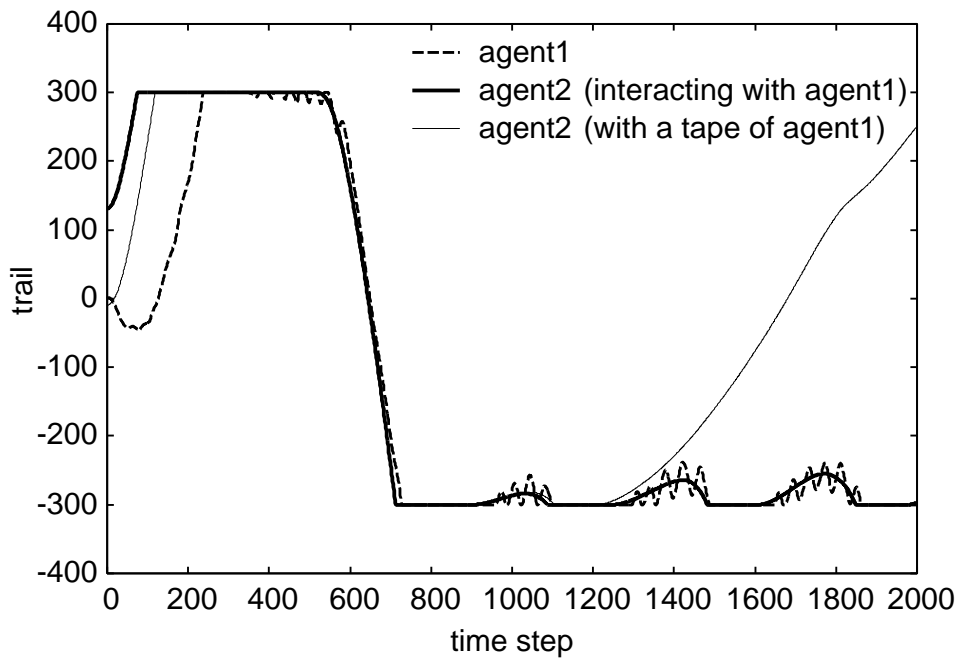


Figure 5.

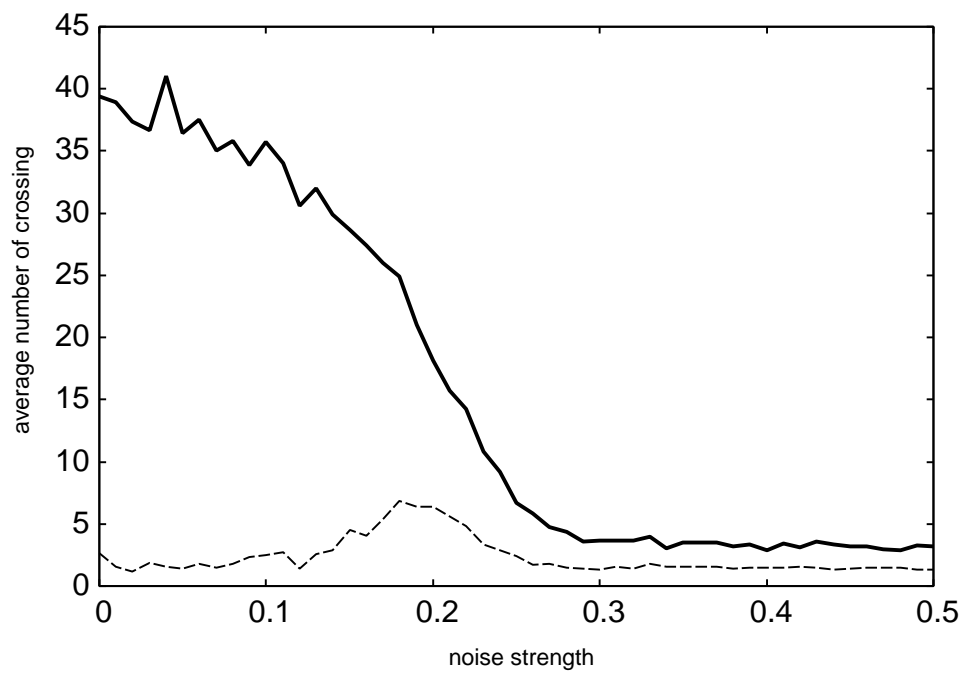


Figure 6.