In Daniel M. Dubois (ed) *Proceedings of CASYS'98, The Second International Conference on Computing Anticipatory Systems.* (New York: Springer-Verlag, 1999).

Autonomy in Anticipatory Systems: Significance for Functionality, Intentionality and Meaning

John D. Collier

Department of Philosophy, University of Newcastle,
University Drive, Callaghan, NSW 2308, Australia
fax: +61 2 4921 7247, email: pljdc@alinga.newcastle.edu.au

Abstract Many anticipatory systems cannot in themselves act meaningfully or represent intentionally. This stems largely from the derivative nature of their functionality. All current artificial control systems, and many living systems such as organs and cellular parts of organisms derive any intentionality they might have from their designers or possessors. Derivative functionality requires reference to some external autonomously functional system, and derivative intentionality similarly requires reference to an external autonomous intentional system. The importance of autonomy can be summed up in the following slogan: No meaning without intention; no intention without function; no function without autonomy. This paper develops the role of autonomy to show how learning new tasks is facilitated by autonomy, and further by representational capacities that are functional for autonomy.

Keywords: autonomy, anticipatory systems, function, intentionality, meaning

INTRODUCTION

Many anticipatory systems cannot in themselves act meaningfully or represent intentionally. This stems largely from the derivative nature of their functionality. All current artificial control systems, and many living systems such as organs and cellular parts of organisms derive any intentionality they might have from their designers or possessors. Derivative functionality requires reference to some external autonomously functional system, and derivative intentionality similarly requires reference to an external autonomous intentional system. The importance of autonomy can be summed up in the following slogan: No meaning without intention; no intention without function; no function without autonomy.

A standard debate concerning computing systems concerns whether they can, non-derivatively, represent meaningfully, and what this would require; see (1), (2), (3), (4), (5). Participants in the debate agree that living systems such as ourselves can be intentional. I propose that a central requirement for non-derivative representation is representational autonomy both modeled on and emergent from functional autonomy like that found in living systems. I consider the central features of our intentionality, and then suggest which of these features a computing system must have in order to embody intentionality and act meaningfully in a non-derivative way.

The issue of meaningful representation is directly relevant to the nature of anticipatory systems, since anticipation requires foreseeing or taking into consideration in advance, either implicitly or explicitly. A non-intentional device might be able to perform anticipative services, but cannot anticipate on its own, since it cannot foresee. Any foresight it has is derivative from its design, and will be consequently design limited. Unless design constraints are kept in mind, the capacities of complex derivatively anticipatory devices are likely to overestimated. However, if the design constraints *are* kept in mind, it is unnecessary to regard derivative devices as anticipatory, but merely as functioning according to the expectations of the designer. On the other hand, if the anticipatory capacities of a device are grounded in its autonomy; cf. (6), (7), (8) and (9), anticipatory design limitations can be overcome through reconsideration of their contribution to autonomy, permitting some new functions to arise which contribute to autonomy in fundamentally new ways. Similar considerations apply to design goals of derivatively autonomous systems, except that they cannot redesign their fundamental anticipatory functions; at best they merely reorganize their design at less fundamental levels to redirect their

anticipatory functions to new derived goals.

AUTONOMY

A system is autonomous if it uses its own information to modify itself and its environment to enhance its survival, responding to both environmental and internal stimuli to modify its basic functions to increase its viability. A major constraint on the survival of an artifact is that it serves its designed purpose: A household robot that makes messes will not last long. Similarly an organism will not last long if its functioning does not contribute well to its autonomy; it will be selected against by natural selection. This inverts the currently popular etiological accounts of function, according to which a function's purpose is that for which it is selected; see (10), (11) and (12), for criticism, see (13) and (14) . The basic idea of the etiological view is that a property P is selected because it does F, and because it does F the organism that possesses P is selected. Instead, on the autonomy view, the autonomy found in the organization of some things (especially organisms and lineages of organisms) which includes F among its functions contributing to autonomy sustains their viability and likelihood of being selected. Thus the basic function of P is to contribute to autonomy, which in turn makes the organism (or its lineage) more viable than it would be without P, all other things being equal. Selection is a result of functionality on this account, not its cause. The standard account focuses on the results, much like behaviorism, rather than the internal causes. In most cases in evolution, function can be understood only with respect to preservation of the lineage rather than the individual organism. This involves history, giving some basis for the etiological view, but the focus should remain on the dynamical process of preservation, not merely the results. Similar considerations apply to lineages of cognitive function in learning processes. Below, I will argue that the autonomy account also applies to lineages of meme transmission.

The self-referential and open character of the problem of self-preservation requires that an autonomous system be flexible, open to signals, capable of self-modification of at least a wide range of its anticipatory functions, and capable of evaluating such modifications. A useful, but not necessary characteristic is the second order capacity to anticipatively self-modify, permitting self driven adaptability. This higher order property requires some sort of self representation and recognition of what could contribute to autonomy, since it is not directly subject to natural selection to weed out unsuccessful modifications. The self-modifications would not have a direct genetic basis (i.e., any differences would have a common genetic context), so selection would not preserve them unless the modifications somehow became genetically fixed. Internal (vicarious) selection may play an important role, perhaps guided by external positive and negative stimuli, but the modifications are basically self-guided.[1]

Naturally autonomous systems have a dynamical (causal) cohesion (15) that is actively maintained by internal and external processes of various kinds that are controlled by their internal information, i.e. they are substantially dynamically self-maintaining (9). The parts of a thing are unified by their unity relation, whose logical closure makes parts of a thing parts of the same thing. Cohesion is the closure of the unity relation among parts of a natural system comprised by the dynamical (including functional) processes that maintain system integrity in the face of external and internal fluctuations. Cohesion thus implies a partition of systems: it individuates systems by the internal cohesion being generally stronger than internal fluctuations and external insults. For example, in a predator both active and passive cohesion (e.g., the largely active food searching behavior followed by eating and metabolism vs. the relatively passive structure of the bones) help to maintain system integrity while at the same time serving to differentiate predators: no two predators interact physiologically or metabolically with each other more than they interact in these ways with themselves.

Autonomous systems have many functional properties that preserve system properties through cycles of interaction, both internally and with the environment. These cycles are typically complex and self-reinforcing. *Process closure* concerns the fact that an overall process must achieve self-reinforcement by supporting system viability, and hence the continuing system capacity to carry out that process. If the system is to achieve overall process closure the elements of the system must interact with each other and with the environment in particular, circumscribed ways. This is *interaction closure*. It is essential to self-regulation, and distinguishes autonomous systems from other cohesive systems like rocks that maintain their integrity merely through

[1)] I ignore the possibility of an organism or device with such a propensity for pathological self-modification that it has little chance for survival. I assume that vicarious selection will eliminate hair-brained schemes; see (6) and (7).

strong bonds that tend to isolate them from other systems, and from systems like gases, and liquids that are more open than solids, but do not have any closure of environmental interaction required for self-regulation; they remain independent only at the whim of environmental contingencies. Although open-ended interaction with the environment makes autonomy a property at the ecological level, in the sense that the closure conditions for its definition make essential reference to the environment, autonomy "belongs" to the autonomous individual in the sense that what makes the difference to autonomy (the organized information controlling cohesion) largely lies in the individual.

Autonomous system processes will in general interact with many other such processes. For example, eating and digesting, etc. support not just hunting capacity that can lead to further eating and digestion, but every other system capacity as well. To maintain themselves, autonomous systems must display a corresponding internal coherency of processes; namely the processes must interrelate flexibly so as to preserve the whole organized complexity that underwrites control of that very responsiveness and adaptability. The functional properties must be so integrated that autonomous systems can maintain an active independence. Unlike all other kinds of systems, autonomous systems are dominated by the organization of these global functional constraints.

From these requirements, it can be seen that autonomy is multidimensional and varies in degree. If the dimensions are distinct enough, we can talk of kinds of autonomy, such as material autonomy, psychological autonomy, social autonomy, and informational autonomy. These kinds of autonomy arise at different levels, and in differing hierarchies of levels, so autonomy is also relative to level and hierarchy. Something that might not be autonomous at the most fundamental physical level, under the extreme conditions found in physics, might be autonomous biologically in the less intense environments of organisms. For example, biological information in hereditary processes depends on metabolic process and environmental interactions, but is grounded in lower level DNA and other macromolecular processes and other physically transmitted processes, but metabolic function itself is grounded in a wider set of molecular processes. Minds, to take another example, might be autonomous in terms of information content, even though they depend on their biological embodiment. In each case, a higher level autonomy will require the existence of an underlying autonomous system, and a kind or level of autonomy will usually contribute to the autonomy of its constituting and embedding level; however, levels and kinds of autonomy can compete just like autonomous individuals compete at the same level. Autonomy may be largely in one dimension or interdependent range of dimensions, despite large dependencies of other kinds. It might be argued, for example, that minds, although highly dependent on bodies materially, are informationally quite autonomous, as would be other autopoietic entities (16). The use of the body by the mind to maintain itself, it might be argued, is analogous to the use of the environment by the body to maintain itself, creating not only an informational independence but arguably making it self-sustaining as well. This maintenance can conflict with bodily function in extreme cases, such as when a hero satisfies his self conception of his personal integrity, and sacrifices his life for unrelated others. If representations can be autonomous, their autonomy will be of this informational kind: they must actively use their own information to maintain their own informational structure and reproduction. Fundamentally, though, they must function to preserve the autonomy of those who have the representations; a wholly "selfish" meme or crazy idea or ideology would soon disappear, if only because its possessors would not survive.

An autonomous device in itself is not especially useful; on the contrary, its behavior on training may be "perverse", since it will respond to enhance its autonomy (and often not well, at that), not the design goals of its training. This can be seen in training natural systems like animals, children and ecologies, in which our best efforts to control the system often leads to the opposite effect. The same frustrations should be expected in training an autonomous robot. Part of the problem is that merely autonomous systems are likely to have minimal anticipatory capacity unless they have sophisticated representational capacity. This is true, for example of young children and ecosystems, which lack the representational capacity to convert verbal ideas or demonstrations into internally governed practices. In such cases, modification of the system behavior requires the simultaneous formation and integration of the required structures and dynamical relations to achieve the desired end. This requires both patience and a good understanding of the system. The trainer has an idea what he or she wants the trainee to learn, but this may not by easy to make compatible with the requirements of autonomy of the trainee. For example, forcing children to eat is usually counterproductive, creating strife at meals. Alternatively, a hungry child will eat some food it does not like, and may come to like the food if it becomes associated and integrated with pleasant experiences. Forcing eating just associates food with

unpleasant experiences, and further integrates resistance to eating into the child's autonomy. The point is that the child can maintain its autonomy with a choice, by refusing to eat and thereby maintaining control of the situation, or control can be removed as an issue, and the child can learn on its own proper eating habits. With the varieties of maintenance of autonomy in mind, the forcing route is likely to be counterproductive.

Autonomy is always self generating, or an autonomous system would not be able to maintain itself. Furthermore, autonomous systems are best formed spontaneously through the integration (through functional and structural cohesion) of their prior properties, altering those properties so that they are constrained by the newly formed cohesion (in other words, they are emergent). They could, in principle, be designed, but the risk is that constraints will be built in that produce a device limited by conscious or unconscious design constraints that prevent spontaneous self-organization, both through restrictions on interactions with the environment, and on internal reorganization to respond to unexpected signals and especially of unexpected signal types (17), (18). The best way to produce an autonomous device is to let it grow under the right conditions.[2] Thus the problem of devising a self-modifying anticipatory device that can develop modifications even to what it can recognize and control is more analogous in some respects to horticulture than to mechanical manufacture.

Programming New Functions: The Advantages of Autonomy

Programming of machines is also greatly facilitated by the ability to pass representations and other memes directly.[3] Programming machines with these capacities can be facilitated by a moderately autonomous representational system, in which the representations themselves have some degree of autonomy, so that both the representational system and individual representations have their own proper functions within the machine. Consider programming a robot to do a task. A standard way to do this with robots of derivative function is to "walk" them through the task, which they then repeat on an appropriate stimulus. Behavior can then be modified by selective corrections. However, there is no chance in this case that the robot will work outside its specific programming and reinforcement training. An autonomous robot, on the other hand, will integrate such training into its autonomous function, being capable of modifying its functions to maintain its autonomy. As mentioned above, this is likely to lead to unacceptable unpredictable results. Deviant robots like this can be selected out or retrained, but the process is likely to take a long time unless the task is very simple and well defined. It would be better to design in representational autonomy so that memes can be transmitted reliably as cohesive and self-maintaining wholes. This is possible, however, only if the representational system itself is functional with respect to the autonomy of its subject, which in turn requires functional integration with the autonomy of the subject. These requirements can ensure that programming will be fully integrated with the anticipative capacities of the device, permitting a high degree of self-control over anticipation, while maintaining the integrity of the initial transmitted meme and designed meme function. Corrections are then possible for general cases rather than specific cases, as with derivative anticipation, and learning is faster. The system would be more like an apprentice than a programmed computer.

The ideal robot is rather like a bright apprentice who can watch the actions of the master, and then copy them in his own way, integrating previously unknown patterns into his actions as best he can. The apprentice starts off clumsily, but with some correction and more demonstration from his master, and considerable practice, his skills improve. They are then integrated into his capacities, and can be used to permit the development of other capacities.[4]

---

2) Cariani (18) gives an example of a physical device invented by Pask in the 1950's that could modify its electrical and electrochemical ferrous sulphate substrate to distinguish tones and magnetic fields. These capacities formed spontaneously through changes in the malleable ferrous sulphate substrate under exposure to appropriate stimuli, rather than being designed in from the beginning. Unfortunately, this work was not pursued further, perhaps due to the dominant computational model of mind.

3) By memes, I refer to transmittable behaviors and the organizational complexes generating these behaviors, without regard to the debate concerning whether or not memes involve anything like the localizable generating structures usually attributed to genes. I include practices, ideologies and paradigms as memes, as well as the more common ideas. I believe that practices are the more fundamental transmitted units.

4) The model is similar to Piaget's model of assimilation, which corresponds to this case in

The role of the master is demonstration, encouragement and correction. Too much of the latter can be counterproductive, first because the apprentice will be more concerned with avoiding errors than integrating practices (contrary to the master's intent), and second because the mistakes of the master are more likely to be passed on. The apprentice's autonomy is enhanced by pleasing the master, and avoiding discipline can supercede actually learning the practice. Individual practice with some idea of what is to be achieved is much more productive than learning from instructions.

REPRESENTATION

The remainder of this chapter discusses the requirements for representational autonomy, including that of memes. According to the pragmatic theory of meaning due to C.S. Peirce (19) the cognitive (intellectual) meaning of a representation is given by our expectations involving its object. We can say, then, that the content of a representation or idea is the information common to all these expectations (some abstraction to more general properties may be implied by the process of isolating the common information), and that this content provides the information needed to reason with the idea. Note that the idea must be integrated with behavior and potential behavior in order for the expectations to exist, so interpretation takes place on a background of pragmatic concerns. The information content of the idea need not exhaust the information content of its physical embodiment, permitting unexpected consequences of representations, even when they do not guide anything but verbal behavior or thought. The revisability of pragmatic meaning in the face of experience requires that, although full articulation might be an ideal, attaining exact correspondence between the information content and the information in the representation's physical embodiment, called "digital information" by Dretske (20), would reduce the ideas to tautologies, with no capacity for producing novel ideas.[5]

Pragmatic ideas themselves need not be autonomous, but they must be functional within an autonomous system in order to have a pragmatic meaning. Because expectations can be falsified both by experience and unexpected consequences of information in representations that are not recognized in the content, pragmatic meaning is somewhat open-ended and revisable much like other functions supporting autonomy. New meanings can either be assembled or rearranged from old ideas or through old forms incorporating new sensory, verbal and proprioceptive information that get incorporated in the preexisting forms to form new representations (this is possible with digital representations as well). Truly novel representations, however, must be able to form spontaneously, so that their organization is self-maintained, and generated by organizing forces driven by a dynamical statistical gradient (this would be thermodynamic in the case of energy alone, but in the case where information is primarily concerned it is perhaps best termed *morphodynamic*. In the case of ideas, if a space of possible (undefined and unarticulated) ideas is larger than the current ones (perhaps the space of ideas is opened by unexpected occurrences), ideas reorganize within this space, perhaps by some Gestalt process, but the details are not well known.[6] The significant thing about genuine creativity is its spontaneous and self-organizing character, permitted by the openness and revisability of pragmatic meaning. It is central to the process of teaching someone (or a suitable device) something they don't know, or even have the immediate capacity for. The high level information to be transmitted that is usually involved in a skill or abstract idea not only needs to be duplicated, but it must be integrated into the functioning of the learning system. This means that ultimately it must be able to contribute to the autonomy of that system, so that it can be used in novel ways, rather than just in the form in which it was transmitted. This is an oft cited but difficult to achieve goal of pedagogy.

Autonomous systems are capable of generating novel functions, including representations in systems including representation systems, on their own, but guidance and the use of tried and true ideas are more efficient. Certain memes are one kind of tested idea that are especially suited for this task, since memes have the capacity to integrate relatively easily in those exposed to them. (Of course this can be a

---

copying the master, and accommodation, which corresponds in this case to integration of the new skills into coordinated practice as well as into a broader range of capacities. The latter allows the apprentice to deal with unexpected eventualities (say the failure of a tool) without needing to rely on the master.

[5] It is worth noting that analytic accounts of meaning tend to assume digitallity.

[6] "Gestalt switches", say in the famous "wife/mother-in-law" figure, take place at the same level, but the formation of new gestalts might introduce a new level of pattern.

disadvantage in the case of unhealthy memes like smoking tobacco or bigotry.) Like genes, meme kinds are transmitted, not meme instances. The autonomy, self-preservation and fidelity of meme transmission are therefore parasitic on meme instance autonomy, which requires representational autonomy within a representational system of an organism or device. This in turn requires some degree of autonomy of the representational system, which must be functionally integrated overall within an autonomous system. The same general requirements hold for derived anticipation, but in this case full analysis of meme transmission and function to the non-autonomous system must include the designing system within which meme instances function to maintain autonomy. In this case the meme instances in the designed artifact itself act merely as constraints on the system's behavioral repertoire, and lack any creative potential (except for bugs and other failures that are more likely to cause problems than resolve them).

Memes are characterized by their capacity to be adopted either consciously or unconsciously, or through some combination of both (as probably happens with the spread of teenage fashions) through their capacity to integrate themselves with the autonomy of their host. This capacity will depend on the functions and structure making up the autonomy of the host, in particular its capacity to reorganize itself so as to integrate the meme rather than merely paying lip service to it (contrast a died in the wool Nazi with a German who merely mimics Nazi rituals to avoid problems). Memes (which include any ideas, practices, ideologies and paradigms) must be learned without previously having the information in the memes, so they must possess some autonomy as instances in order to perform the organization required to integrate themselves into their hosts' functionality. This autonomy need not be large; most of the integrating capacity will be in the host, and the meme must serve some function for the host or it will be quickly rejected, but there must be some capacity in the information contained in the meme to guide this integration.

Returning to the apprentice example, whatever skill is being taught must be something that humans can learn to do with minimal guidance but much practice, leading to mastery of the skill. Mastering the skill implies not only the ability to apply it to unforseen circumstances, but being able to develop new skills to achieve related but unforseen goals, i.e., the capacity not only to anticipate, but to develop new anticipative skills, and perhaps even to anticipate what skills might be eventually required. In fact, the last is required in order to create the disequilibrium required to spontaneously develop novel skills. Ideally, anticipatory devices should have the same capacities in order both to be trained in such a way as to be able to deal with new situations, and with new kinds of situations. This may be a disadvantage if the task and all its parameters are well defined, in which case a well-designed single purpose machine may serve best; however, a general purpose robot requires autonomy first, and the capacity to either copy or follow instructions in accord with its autonomy, and to develop the ability to recognize regular failures and be able to create and test novel solutions. Essentially, this would be a robot susceptible to meme hosting.

The distinction between autonomous anticipation and derivative anticipation might seem to be of only theoretical interest, given our present design abilities and the capacities of current computers. The distinction should be kept in mind, however, or else the capacities of systems with the only derivative anticipation are likely to be overestimated, and those of systems with the latter (our goal for modeling) to be underestimated. This can lead to overly optimistic claims for derivative systems, and disappointing results that might harm the whole project of developing anticipative systems, derivative or not. In the longer run, experiments with self-maintaining robots and the development of autonomous machines may lead to trainable robots. This work should be integrated with the neglected work by Pask (18) on electrochemically interconnected malleable substrates. Purely mechanical designs seem to lack the capacity for self-organization required for true autonomy and non-derivative function, including adaptive and anticipatory capacities.

CONCLUSION

Autonomy is fundamental to meaningful action and representation, and thus to anticipation. An anticipatory system may function either derivatively or autonomously. Only in the latter case is independent anticipation possible. Autonomy creates certain difficulties for training, but it permits a system to integrate its learning into the organization that constitutes its autonomy, allowing further development, improvement, and applications to novel cases. Derivative anticipation, on the other hand, is limited to pre-defined circumstances, and can function only in the service of other, autonomous systems.

Transmissible practices, ideas and such (memes) must have a certain degree of autonomy to allow their transmission and integration into the autonomy of new systems. Computing anticipatory systems that can learn from other systems through training by

being exposed to memes must be autonomous to perform this function. This requirement has not been generally been taken into consideration in past work in artificial intelligence or in the design of anticipatory systems. Ignoring this requirement can lead to overestimating the capacities of designed systems, and underestimating the difficulties involved in designing self controlling anticipatory systems.

ACKNOWLEDGMENTS

Thanks are owed to the Newcastle Dynamic Organized Complex Adaptive Systems Group. I was slow to accept the non-etiological account of function, despite the best efforts of Mark Bickhard and Wayne Christensen. The central idea of autonomy was developed mutually by Wayne Christensen, Cliff Hooker and myself, but Wayne made the largest contribution. My application of the concept in this paper is perhaps more free than either would allow.

REFERENCES

1. Dennett, D.C., *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
2. Dretske, F., "Machines and the Mental," *Proceedings and Addresses of the APA* **59**: 23-33 (1985).
3. Searle, J., "Minds, Brains and Programs," *Behavioral and Brain Sciences* 3,: 417-58 (1980).
4. Searle, J., *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992.
5. Churchland, P.F. (1996). *The Engine of Reason, The Seat of the Soul*. Cambridge, MA: MIT Press.
6. Bickhard, M.H., "Representational Content in Humans and Machines," *Experimental and Theoretical Artificial Intelligence* **5**, 285-333 (1993).
7. Bickhard, M. H. and Terveen, L.. *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. New York: Elsevier, 1995.
8. Christensen, W.D., "A Complex Systems Theory of Teleology," *Biology and Philosophy* **11**: 301-320 (1996).
9. Christensen, W.D., Collier, J.D. and Hooker, C.A., "Adaptiveness and Adaptation: A New Autonomy-theoretic Analysis and Critique," *Biology and Philosophy* (submitted).
10. Wright, L., "Functions," *Philosophical Review* **82**, 139-168 (1973).
11. Millikan, R.G., "In Defense of Proper Functions," *Philosophy of Science* **56**, 288-302 (1989).
12. Neander, K., "Functions As Selected Effects: The Conceptual Analyst's Defense," *Philosophy of Science*, **58**, 168-184 (1991).
13. Christensen, W.D. and Hooker, C.A., "Autonomous Systems and Self-Directed Heuristic Policies: Toward New Foundations for Intelligent Systems," Hayes, B., Heath, R. , Heathcote A., and Hooker, C.A. (eds), *Proceedings of the Fourth Australian Cognitive Science Conference*, Newcastle, Australia, 1997.
14. Foss, J., "On the Evolution of Intentionality as Seen from the Intentional Stance," *Inquiry* **37**: 287-310 (1994).
15. Collier, J. D., "Supervenience and reduction in biological hierarchies," M. Matthen and B. Linsky (eds) *Philosophy and Biology: Canadian Journal of Philosophy Supplementary Volume* 14, 209-234 (1988).
16. Maturana, H and Varella, F., *Autopoiesis and Cognition*. Dordrecht: Reidel, 1980.
17. Carani, P., "Emergence and Artificial Life', C.G. Langton, J.D. Farmer and S. Rasmussen (eds) *Artificial Life II, SFI Studies in Complexity, vol 10*. Addison-Wesley, 1991, 775-797.
18. Carani, P., "To Evolve and Ear," *Systems Research* **10**, 19-33 (1993).
19. Peirce, C.S., *Collected Papers*, edited by Charles Hartshorne and Paul Weiss, Cambridge University Press, 1960.
20. Dretske, F., *Knowledge and the Flow of Information*. MIT Press, 1981.