# Robust Explanations in Artificial Life

Eric Silverman[1] and Takashi Ikegami[1]

[1]Department of General Systems Studies,
The Graduate School of Arts and Sciences,
The University of Tokyo, 3-8-1 Komaba, Tokyo 153-8902
erics@sacral.c.u-tokyo.ac.jp
ikeg@sacral.c.u-tokyo.ac.jp

## Extended Abstract

Finding robust explanations of behaviours in Alife and related fields is made difficult by the lack of any formalised definition of robustness. A concerted effort to develop a framework which allows for robust explanations of those behaviours to be developed is needed, as well as a discussion of what constitutes a potentially useful definition for behavioural robustness. To this end, we must differentiate between two senses of robustness: robustness in systems; and robustness in explanation.

When discussing systems, robustness is often described as a property which gives the system a certain resilience against perturbation. A robust system is thus able to retain functionality despite variation. In contrast, we define a robust explanation as a scientific explanation which can identify causal factors that underlie a phenomenon in a variety of circumstances.

The concept of robustness analysis, pioneered by Levins (1966), has illuminated the importance of developing a comprehensive research programme to develop such explanations. Levins argues that doing so requires the study of multiple models of that same phenomenon. Each model should be distinct, containing differing core assumptions or methodologies. If these different models still produce similar results, we can develop what Levins calls a robust theorem: an explanation of the behaviour of interest which is largely independent of the details of the models being studied.

The difficulty for Alife researchers lies in developing an appropriate set of models to produce robust explanations. Weisberg (2005) provides an intensive examination of robustness analysis, describing the concept of a robust property, or a property common to multiple models which contain different idealising assumptions. This leads to a discussion of the need to find common structures between models: those elements which give rise to the robust property. However, many models in Alife not only have different idealising assumptions, but may be based on vastly different methodologies entirely.

In order to escape this conundrum, we need a unified framework under which to search for common structures in order to perform robustness analysis. Models in Alife can frequently share a conceptual relationship - they examine similar behaviours within biological systems, but using fundamentally different methods. The way forward is to create experiments and simulations which share common grounding and related contexts, even when these experiments are quite different in implementation.

An examination of our own work in robotics (Hubert et al, 2009) and biochemical experiments (Ikegami 2009) will provide an example of how divergent methodologies can be used to develop a framework of idealising assumptions. This framework can then form the basis for the development of robust explanations. The commonalities found between the robust behavior of the robot (Hubert et al, 2009) and the biochemical experiments (Ikegami 2009) demonstrate recovery mechanisms which can keep a system from degrading into non-moving states. Here self-movement creates robustness and robustness enables ”intentional” behavior. Through an examination of these common structures, we can begin to develop a framework for robust explanations of these self-movement behaviours.

## References

Hubert, J., Matsuda, E., Silverman, E., and Ikegami, T. (2009). A robotic approach to understanding robustness. In Proceedings of Mobiligence 2009, Awaji, Japan.

Ikegami, T. (2009). The search for a first cell under the maximalism design principle. Journal of Technoetic Arts, In Press.

Levins, R. (1966). The strategy of model-building in population biology. American Scientist, 54:421–431.

Weisberg, M. (2005). Robustness analysis. Philosophy of Science, 73:730–742.