

Version 4.1

Expanding the modern synthesis II: Formal perspectives on the inherent role of niche construction in fitness

Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute*

October 20, 2010

Abstract

Expanding the modern synthesis requires elevating the role of interaction within and across various biological scales to the status of an evolutionary principle. One way to do this is to characterize genes, gene expression, and embedding environment as information sources linked by crosstalk, constrained by the asymptotic limit theorems of information theory (Wallace, 2010a). This produces an inherently interactive structure that escapes the straightjacket of mathematical population genetics and other replicator dynamics. Here we examine fitness from that larger perspective, finding it intimately intertwined with niche construction. Two complementary models are explored: niche construction as mediating the connection between environmental signals and gene expression, and as a means of tuning the channel for the transmission of genetic information in a noisy environment. These are different views of the same elephant, in a sense, seen as simplified projections down from a larger dynamic system.

Key Words: average distortion, catalysis, evolution, information theory.

1 Introduction

Lewontin (2010), reviewing the book by Fodor and Piatelli-Palmarini (2010), describes the modern evolutionary synthesis as having four basic metaphorical principles:

(1) **The principle of variation:** Among individuals in a population there is variation in form, physiology, and behavior.

*Contact: Rodrick Wallace, PhD, Box 47, NYSPI, 1051 Riverside Drive, New York, NY, 10032, rodrick.wallace@gmail.com

(2) **The principle of heredity:** Offspring resemble their parents more than they resemble unrelated individuals.

(3) **The principle of differential reproduction:** In a given environment, some forms are more likely to survive and produce more offspring than other forms.

(4) **The principle of mutation:** New heritable variation is constantly occurring.

Lewontin, however, finds this structure lacking, in that “...there is an immense amount of biology that is missing”, largely involving interactions within and across structural entities at various scales.

To address this lack, Wallace (2010a) introduces a fifth principle:

(5) **The principle of environmental interaction:** Individuals and groups engage in powerful, often punctuated, dynamic mutual relations with their embedding environments that may include the exchange of heritage material between markedly different organisms.

The central innovation of that work is to avoid the intellectual straightjacket of mathematical population genetics and other replicator dynamics formalisms by describing embedding ecosystem, genetic heritage, and (cognitive) gene expression in terms of interacting information sources whose dynamics are driven by the homology between information source uncertainty and free energy density in a series of regression-like relations similar to the empirical Onsager equations of nonequilibrium thermodynamics. Taking much the perspective of Champagnat et al. (2006), the resulting theory is inherently coevolutionary, in the largest sense, so that there is no single ‘natural’ scale at which ‘selection’ takes place. There is, rather, a set of interactive quasi-equilibria subject to often highly structured large deviations representing the punctuated equilibria of Eldredge and Gould (1972), as well as possibly larger, hierarchical, conformations.

Here we restrict the theory to recover a closer analog to conventional selection and fitness theory, inherently modified, however, by processes of niche construction (e.g., Odling-Smee et al., 2003) such as biofilms, multicellularity, burrows, eusocial nests, and larger social formations that mediate between individual development and environmental signals. This is a necessary test of the larger theory, in the same logical sense that any modern theory of gravitation must reduce to Newton’s in an appropriate approximation.

We examine two complementary perspectives, first, in terms of a channel connecting gene expression to environmental signals, and then in terms of the transmission of genetic information along a noisy environmental channel. These are, in a sense, two downward projected views of the more complete elephant described in Wallace (2010a).

2 Niche construction as lubricating the channel connecting environmental signals to gene expression

The basic schema is that of figure 1.

A multifactorial environmental signal, a ‘message’, $y^n = \{y_1, y_2, \dots, y_n\}$, representing the systematic output of an embedding ecosystem information source, is expressed by a cognitive gene expression information source – following Wallace (2010a) and Wallace and Wallace (2009) – in terms of a multifactorial pattern of phenotypes, the ‘message’ $b^n = \{b_1, b_2, \dots, b_n\}$. We suppose it possible to deterministically retranslate, ‘decode’, the phenotype message to produce a new version of the original environmental message, i.e., the environment inferred from the phenotype. We write that inferred picture as $\hat{y}^n = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$. We then introduce a numerical distortion measure that compares y_i with \hat{y}_i , writing it as $d(y_i, \hat{y}_i)$. As Cover and Thomas (1991) indicate, many such measures are possible and have been used, and the essential dynamics are, remarkably, independent of the precise measure chosen.

Suppose that with each path y^n and b^n -path retranslation into the y language, \hat{y}^n , there are associated individual, joint, and conditional probability distributions $p(y^n), p(\hat{y}^n), p(y^n, \hat{y}^n), p(y^n | \hat{y}^n)$.

The average distortion is defined as

$$D = \sum_{y^n} p(y^n) d(y^n, \hat{y}^n).$$

(1)

Clearly, D is an inverse fitness measure: phenotypes that match environments produce higher rates of successful reproduction, in this model.

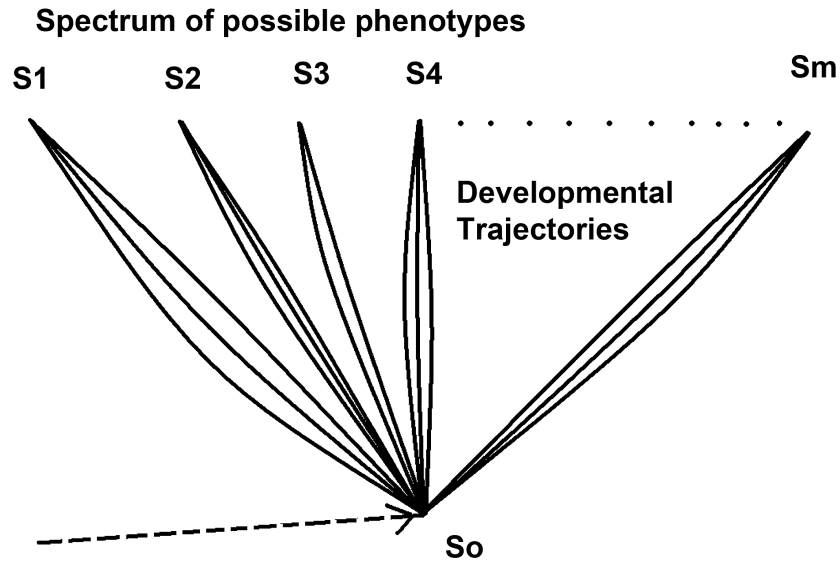
It is possible, using the distributions above, to define the information transmitted from the Y to the \hat{Y} process using the Shannon source uncertainty of the strings:

$$I(Y, \hat{Y}) = H(Y) - H(Y|\hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y}),$$

(2)

where $H(\dots, \dots)$ is the standard joint, and $H(\dots|\dots)$ the conditional, Shannon uncertainties (Cover and Thomas, 1991, Ash, 1990). If there is no uncertainty

Multifactorial phenotype expression of environmental message
{b1, b2, ..., bn}



Multifactorial environmental message
{y1, y2, ..., yn}

Figure 1: Environmental ‘message’ $y = \{y_1, y_2, \dots, y_n\}$ is expressed as the phenotype ‘message’ $b = \{b_1, b_2, \dots, b_n\}$, that is deterministically translated into a reconstruction of the environmental message, $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$. A distortion measure $d(y, \hat{y})$ characterizes the difference between what was sent and what was received, allowing construction of an average distortion D , inversely measuring fitness, and a rate distortion function $R(D)$ that, as a free energy analog, permits analysis in terms of available metabolic free energy, κM .

in Y given the retranslation \hat{Y} , then no information has been lost, and the systems are in perfect synchrony. In general, of course, this will not be true, and information has been lost in the translation of the environmental message into final phenotype spectrum.

The *rate distortion function* $R(D)$ for a source Y with a distortion measure $d(y, \hat{y})$ is defined as

$$R(D) = \min_{p(y, \hat{y}); \sum_{(y, \hat{y})} p(y)p(\hat{y})d(y, \hat{y}) \leq D} I(Y, \hat{Y}). \quad (3)$$

The minimization is over all conditional distributions $p(y|\hat{y})$ for which the joint distribution $p(y, \hat{y}) = p(y)p(y|\hat{y})$ satisfies the average distortion constraint (i.e., average distortion $\leq D$).

The *Rate Distortion Theorem* states that $R(D)$ is the minimum necessary rate of information transmission that ensures the communication between the biological vesicles does not exceed average distortion D . Thus $R(D)$ defines a minimum necessary channel capacity. Cover and Thomas (1991) or Dembo and Zeitouni (1998) provide details. The rate distortion function has been calculated for a number of systems.

An absolutely central fact characterizes the rate distortion function: Cover and Thomas (1991, Lemma 13.4.1) show that $R(D)$ is necessarily a decreasing convex function of D for any reasonable definition of distortion.

That is, $R(D)$ is always a reverse J-shaped curve. This will prove crucial for the overall argument. Indeed, convexity is an exceedingly powerful mathematical condition, and permits deep inference (e.g., Rockafellar, 1970). Ellis (1985, Ch. VI) applies convexity theory to conventional statistical mechanics.

For a Gaussian channel having noise with zero mean and variance σ^2 (Cover and Thomas, 1991),

$$R(D) = 1/2 \log[\sigma^2/D], 0 < D \leq \sigma^2$$

$$R(D) = 0, D > \sigma^2. \quad (4)$$

For a Poisson channel with message arrival rate λ – essentially information transmission by an analog to photon counting in which the instantaneous output

rate is controlled by the input modulo noise – the rate distortion function is (Bedekar, 2001)

$$R(D) = \log[1/\lambda D], 0 < D \leq 1/\lambda$$

$$R(D) = 0, D > 1/\lambda$$

(5)

where D is the average in-order service time of a hypothetical first come first serve queue that would result in the output.

Recall, now, the relation between information source uncertainty and channel capacity (e.g., Ash, 1990), $H[X] \leq C$, where H is the uncertainty of the source X and C the channel capacity, defined according to the relation (Ash, 1990)

$$C \equiv \max_{P(X)} I(X|Y),$$

(6)

where $P(X)$ is chosen so as to maximize the rate of information transmission along a channel Y .

Finally, recall the analogous definition of the rate distortion function above, again an extremum over a probability distribution.

Our own work (e.g., Wallace and Wallace, 2008, 2009) focuses on the homology between information source uncertainty and free energy density. More formally, if $N(n)$ is the number of high probability ‘meaningful’ – that is, grammatical and syntactical – sequences of length n emitted by an information source X , then, according to the Shannon-McMillan Theorem, the zero-error limit of the Rate Distortion Theorem (Ash, 1990; Cover and Thomas, 1991; Khinchin, 1957),

$$\begin{aligned} H[X] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \\ &= \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) \end{aligned}$$

$$= \lim_{n \rightarrow \infty} \frac{H(X_0, \dots, X_n)}{n+1},$$

(7)

where, again, $H(\dots|\dots)$ is the conditional and $H(\dots, \dots)$ is the joint Shannon uncertainty.

In the limit of large n , $H[X]$ becomes homologous to the free energy density of a physical system at the thermodynamic limit of infinite volume. More explicitly, the free energy density of a physical system having volume V and partition function $Z(\beta)$ derived from the system's Hamiltonian – the energy function – at inverse temperature β is (e.g., Landau and Lifshitz 2007)

$$F[K] = \lim_{V \rightarrow \infty} -\frac{1}{\beta} \frac{\log[Z(\beta, V)]}{V} \equiv$$

$$\lim_{V \rightarrow \infty} \frac{\log[\hat{Z}(\beta, V)]}{V},$$

(8)

with $\hat{Z} = Z^{-1/\beta}$. The latter expression is formally similar to the first part of equation (7), a circumstance having deep implications: Feynman (2000) describes in great detail how information and free energy have an inherent duality. Feynman, in fact, defines information precisely as the free energy needed to erase a message. The argument is surprisingly direct (e.g., Bennett, 1988), and for very simple systems it is easy to design a small (idealized) machine that turns the information within a message directly into usable work – free energy. Information is a form of free energy and the construction and transmission of information within living things consumes metabolic free energy, with nearly inevitable losses via the second law of thermodynamics. If there are limits on available metabolic free energy there will necessarily be limits on the ability of living things to process information, and in particular to match phenotype expression to environmental cues.

$R(D)$ is, by this account, a free energy measure, constrained by the availability of metabolic free energy. We therefore write a classic probability density function for the average distortion D – an inverse fitness measure – as

$$P[D] = \frac{\exp[-R(D)/\kappa M]}{\int_0^{max} \exp[-R(D)/\kappa M] dD}$$

(9)

where κ represents the efficiency of use of available metabolic energy rate M .

Higher κM , in this model, permits lower average distortion, and we can calculate the mean of the average distortion – again, an inverse index of fitness – as a function of κM .

For the Gaussian channel, direct calculation gives

$$\langle D_G \rangle = \int DP[D]dD = \frac{\sigma^2(2\kappa M + 1)}{4\kappa M + 1}.$$

(10)

For the Poisson channel,

$$\langle D_P \rangle = \frac{\kappa M + 1}{\lambda(2\kappa M + 1)}.$$

(11)

At the limits of high and low κM , these have, respectively, the values

$$\langle D_G \rangle = \frac{\sigma^2}{2}, \sigma^2,$$

$$\langle D_P \rangle = \frac{1}{2\lambda}, \frac{1}{\lambda}.$$

(12)

According to this model, $\langle D \rangle$, which represents an inverse fitness measure, is – within a factor of 2 – primarily driven by internal indices of the way in which a developing organism can represent environmental signals.

This result has significant implications for understanding niche construction: in the sense of a recent treatment of intrinsically disordered proteins in a rough folding funnel (Wallace, 2010b), niche construction ‘self-lubricates’ or impedance-matches the relation between gene expression and environmental demands, decreasing σ^2 or increasing λ , according to the nature of the channel.

3 Niche construction as tuning the channel for the transmission of genetic information in a noisy environment

A quite different picture of niche construction emerges from application of the ‘tuning’ version of Shannon’s Coding Theorem to instantiate a Rate Distortion Manifold (Glazebrook and Wallace, 2009a, b). The basic argument follows that of Wallace et al., (2009, Section 2.10). The focus is now on the successful transmission of genetic heritage within a noisy environmental channel, and the role of niche construction in tuning that channel.

Messages from an information source, seen as symbols x_j from some alphabet, each having probabilities P_j associated with a random variable X , are encoded into the language of some transmission channel, a random variable Y with symbols y_k , having probabilities P_k , possibly with error. The received symbol y_k is then retranslated – decoded – (without error) into some x_k , which may or may not be the same as the x_j that was sent.

More formally, the message sent along the channel is characterized by a random variable X having the distribution $P(X = x_j) = P_j, j = 1, \dots, M$. The channel through which the message is sent is characterized by a second random variable Y having the distribution $P(Y = y_k) = P_k, k = 1, \dots, L$.

The joint probability distribution of X and Y is defined as $P(X = x_j, Y = y_k) = P(x_j, y_k) = P_{j,k}$ and the conditional probability of Y given X is $P(Y = y_k | X = x_j) = P(y_k | x_j)$.

Recall that the Shannon uncertainty of X and Y independently and the joint uncertainty of X and Y together are defined respectively as

$$\begin{aligned} H(X) &= - \sum_{j=1}^M P_j \log(P_j), \\ H(Y) &= - \sum_{k=1}^L P_k \log(P_k), \\ H(X, Y) &= - \sum_{j=1}^M \sum_{k=1}^L P_{j,k} \log(P_{j,k}). \end{aligned}$$

The conditional uncertainty of Y given X is defined as

$$H(Y|X) = - \sum_{j=1}^M \sum_{k=1}^L P_{j,k} \log[P(y_k | x_j)].$$

For any two stochastic variates X and Y , $H(Y) \geq H(Y|X)$, as knowledge of X generally gives some knowledge of Y , and equality occurs only in the case of stochastic independence.

$$\text{Since } P(x_j, y_k) = P(x_j)P(y_k | x_j), H(X|Y) = H(X, Y) - H(Y).$$

The information transmitted by translating the variable X into the channel transmission variable Y – usually with error – and then decoded without error the transmitted Y back into X is defined as $I(X|Y) \equiv H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$.

See, for example, Ash (1990), Khinchin (1957) or Cover and Thomas (1991) for details. The essential point is that if there is no uncertainty in X given the channel Y , then there is no loss of information through transmission. In general this will not be true, and herein lies the essence of the theory.

Given a fixed vocabulary for the transmitted variable X , and a fixed vocabulary and probability distribution for the channel Y , we may vary the probability distribution of X in such a way as to maximize the information sent. Recall that the capacity of the channel is defined according to equation (6), i.e., $C \equiv \max_{P(X)} I(X|Y)$, subject to the subsidiary condition that $\sum P(X) = 1$.

The critical trick of the Shannon Coding Theorem for sending a message with arbitrarily small error along the channel Y at any rate $R < C$ is to encode it in longer and longer ‘typical’ sequences of the variable X ; that is, those sequences whose distribution of symbols approximates the probability distribution $P(X)$ above which maximizes C .

If $S(n)$ is the number of such ‘typical’ sequences of length n , then $\log[S(n)] \approx nH(X)$, where $H(X)$ is the uncertainty of the stochastic variable defined above. Some consideration shows that $S(n)$ is much less than the total number of possible messages of length n . Thus, as $n \rightarrow \infty$, only a vanishingly small fraction of all possible messages is meaningful in this sense. This observation, after some considerable development, is what allows the Coding Theorem to work so well. In sum, the prescription is to encode messages in typical sequences, that are sent at very nearly the capacity of the channel. As the encoded messages become longer and longer, their maximum possible rate of transmission without error approaches channel capacity as a limit. Again, Ash (1990), Khinchin (1957) and Cover and Thomas (1991) provide details.

This approach can be, in a sense, inverted to give a tuning theorem which parsimoniously describes the essence of the Rate Distortion Manifold.

Telephone lines, optical wave, guides and the tenuous plasma through which a planetary probe transmits data to earth may all be viewed in traditional information-theoretic terms as a noisy channel around which we must structure a message so as to attain an optimal error-free transmission rate.

Telephone lines, wave guides, and interplanetary plasmas are, relatively speaking, fixed on the timescale of most messages, as are most other signaling networks. Indeed, the capacity of a channel, is defined by varying the probability distribution of the ‘message’ process X so as to maximize $I(X|Y)$.

Suppose there is some message X so critical that its probability distribution must remain fixed – here, the genetic heritage. The trick is to fix the distribution $P(x)$ but modify the channel – that is, tune it – so as to maximize $I(X|Y)$. Such a *dual* channel capacity C^* can be defined as

$$C^* \equiv \max_{P(Y), P(Y|X)} I(X|Y).$$

(13)

But

$$C^* = \max_{P(Y), P(Y|X)} I(Y|X)$$

since

$$I(X|Y) = H(X) + H(Y) - H(X, Y) = I(Y|X).$$

Thus, in a purely formal mathematical sense, the message transmits the channel, and there will indeed be, according to the Shannon Coding Theorem, a channel distribution $P(Y)$ which maximizes C^* .

One may do better than this, however, by modifying the channel matrix $P(Y|X)$, in our case, via niche construction. Since $P(y_j) = \sum_{i=1}^M P(x_i)P(y_j|x_i)$, $P(Y)$ is entirely defined by the channel matrix $P(Y|X)$ for fixed $P(X)$ and

$$C^* = \max_{P(Y), P(Y|X)} I(Y|X) = \max_{P(Y|X)} I(Y|X).$$

(14)

Calculating C^* requires maximizing the complicated expression $I(X|Y) = H(X) + H(Y) - H(X, Y)$ containing products of terms and their logs, subject to constraints that the sums of probabilities are 1 and each probability is itself between 0 and 1. Maximization is done by varying the channel matrix terms $P(y_j|x_i)$ within the constraints. This is a difficult problem in nonlinear optimization, but for the special case $M = L$, C^* , the answer can be found by inspection. If $M = L$, then choose $P(y_j|x_i) = \delta_{j,i}$, where $\delta_{i,j}$ is 1 if $i = j$ and 0 otherwise. For this special case $C^* \equiv H(X)$, with $P(y_k) = P(x_k)$ for all k . Information is thus transmitted without error when the channel becomes ‘typical’ with respect to the fixed message distribution $P(X)$.

If $M < L$, matters reduce to this case, but for $L < M$ information must be lost, leading to Rate Distortion limitations.

Thus modifying the channel may be a far more efficient means of ensuring transmission of an important message than encoding that message in a ‘natural’ language that maximizes the rate of transmission of information on a fixed channel.

We have examined the two limits in which either the distributions of $P(Y)$ or of $P(X)$ are kept fixed. The first provides the usual Shannon Coding Theorem, and the second a niche construction tuning theorem variant, a tunable retina-like Rate Distortion Manifold. It seems likely, however, than for many important systems $P(X)$ and $P(Y)$ will interpenetrate, so that $P(X)$ and $P(Y)$ will affect each other in characteristic ways, so that some form of mutual tuning may be the most effective strategy, a tuning that can perhaps be permanently recorded via adaptive mutations or learned cultural changes.

The essential point is that, for the transmission of genetic heritage in a noisy environment, niche construction provides a means to tune the transmitting channel around the message.

4 Discussion and conclusions

Taking a gene expression-environment signal perspective, according to equation (12), an organism whose developmental ability to respond to environmental signals, to match phenotype to the demands of an embedding ecosystem, is constrained by channel capacity – high developmental noise σ^2 or low signal rate λ – must inevitably suffer diminished fitness.

But what of an organism that can interact, in a hierarchical manner, with the embedding environment, in the sense of principle (5)? Biofilms, multicellularity, burrows, eusocial nests, herds, and the more structured social assemblages of the hominids and their cultures, all provide means, in this model, to limit noise or increase signal strength, by, essentially, impedance matching environment to development. That is, examining the fitness of the organism in terms of its constructed niche decreases an ‘effective’ σ^2 and increases an ‘effective’ λ in equation (12), representing increased fitness in terms of decreased distortion in the channel between the organism and an embedding environment now primarily represented by the constructed niche.

Another way of looking at this is to say that it is the hierarchical structure of organism-in-constructed niche that interacts with the environment, now via the effective parameters.

Looking at the transmission of genetic heritage in a noisy environment – a quite different view – Section 3 suggests that niche construction can also be viewed as a means of tuning the gene transmission channel so as to maximize the fidelity of the genetic message.

These are, then, complementary perspectives that describe the same process via added epicycles to the limited machinery of the first four principles. Introduction of principle (5) collapses the argument, in the sense of Wallace (2010a), and subsumes both models, providing a more complete picture of the entire elephant, in a manner of speaking, at the expense of the considerable mathematical machinery inherent to equation (4) of that paper.

Introducing animal culture, in the sense of Avital and Jablonka (2000), directly into the model of Wallace (2010a) is most simply done by invoking a fourth information source – culture as a generalized language – into equation

(4) of that work. The argument is direct, although the effects are likely to be most subtle.

Clearly, then, the introduction of principle (5) permits incorporation into formal theory of at least some of the immense amount of biology that is missing from the modern synthesis. Most critically, fitness and niche construction must become intimately intertwined in any such expanded formulation.

5 Acknowledgments

The author thanks D.N. Wallace and E. Boney for useful discussions.

6 References

- R. Ash, *Information Theory*, Dover, New York, 1990.
- E. Avital, E. Jablonka, *Animal Traditions: Behavioral Inheritance in Evolution*, Cambridge University Press, NY, 2000.
- A. Bedekar, On the information about message arrival times required for in-order decoding. In *Proceedings of the International Symposium on Information Theory (ISIT)*, Washington, DC, 2201:227, 2001.
- C. Bennett, Logical depth and physical complexity. In R. Herkin (ed.), *The Universal Turing Machine: A Half-Century Survey*, Oxford University Press, pp. 227-257, 1988.
- N. Champagnat, R. Ferriere, S. Meleard, Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models, *Theoretical Population Biology*, 69(2006) 297-321.
- T. Cover, J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
- A. Dembo, O. Zeitouni, *Large Deviations and Applications*, 2nd. edn., Springer, NY, 1998.
- N. Eldredge, S. Gould, Punctuated equilibrium: an alternative to phyletic gradualism. In T. Schopf (ed.), *Models in Paleobiology*, 82-115, Freeman, Cooper and Co., San Francisco, 1972.
- R. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer, New York, 1985.
- R. Feynman, *Lectures on Computation*, Westview, New York, 2000.
- J.Fodor, M. Piattelli-Palmarini, *What Darwin Got Wrong*, Farrar, Straus, and Giroux, 2010.
- J.F. Glazebrook, R. Wallace, Small worlds and Red Queens in the global workspace: an information theoretic approach, *Cognitive Systems Research*, 10:333-365 (2009a).
- J.F. Glazebrook, R. Wallace, Rate distortion manifolds as models for cognitive information, *Informatica*, 33:309-345 (2009b).
- A. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York, 1957.

- L. Landau and E. Lifshitz, *Statistical Physics, Part I*, Elsevier, New York, 2007.
- R. Lewontin, Not so natural selection, *New York Review of Books Online*, May 27 (2010).
- F. Odling-Smee, K. Laland, M. Feldman, *Niche Construction: The Neglected Process in Evolution*, Princeton University Press, N.J, 2003.
- R. Rockefellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- R. Wallace, D. Wallace, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, *Transactions on Computational Systems Biology IX*, LNBI 5121(2008) 23-85.
- R. Wallace, D. Wallace, Code, context, and epigenetic catalysis in gene expression, *Transactions on Computational Systems Biology XI*, LNBI 5750(2009) 283-334.
- R. Wallace, D. Wallace, *Gene Expression and Its Discontents: The Social Production of Chronic Disease*, Springer, New York, 2010.
- R. Wallace, D. Wallace, R.G. Wallace, *Farming Human Pathogens: Ecological Resilience and Evolutionary Process*, Springer, New York, 2010.
- R. Wallace, *Consciousness: a mathematical treatment of the global neuronal workspace model*, Springer, New York, 2005.
- R. Wallace, Expanding the modern synthesis, *Comptes Rendus Biologies*, doi:10.1016/j.crv.2010.08.003 (2010a).
- R. Wallace, Structure and dynamics of the ‘protein folding code’ inferred using Tlusty’s topological rate distortion approach, *BioSystems*, doi:10.1016/j.biosystems.2010.09.007 (2010b).