

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/111857>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

On a general theoretical foundation for endocrinology

Hugo A. van den Berg

Mathematics Institute — University of Warwick, Coventry CV4 7AL

Email: hugo@maths.warwick.ac.uk

Abstract — Hormones are messenger molecules that distribute across all tissues and thus operate on the whole-organism level. Moreover, a given hormone typically affects a number of different biological processes. As such, hormones coordinate concerted cooperation between the cells and tissues of an organism, a phenomenon that has been termed “organismal harmony.” Furthermore, a concept that recently has been gaining traction is that hormones mediate or represent life-history trade-offs which are ultimately moulded by evolutionary pressures. Here, this concept is extended to include all “decisions” or “choices” that are made at the organismal level. A formal framework is sketched to explore the proposition that organismal biology, together with the “fitness landscape,” suffice in principle to determine a minimalistic dynamics of the endocrine system.

Keywords: Endocrinology, fitness, mathematical biology, evolution of the endocrine system, philosophy of biology

Introduction

Newcomers to endocrinology find themselves faced with the bewildering complexity that arises as each hormone affects not just one, but an entire “suite” of physiological or developmental processes. Moreover, each such process is typically the target of several hormones¹, which obliges the poor student to memorise lists of hormonal actions and their various cross-talking connections. To understand the latter, it may help to visualise them graphically, for instance as stops along underground railway lines, where each line corresponds to a given hormone (Fig. 1). Endocrinologists use the term *pleiotropy* in reference to the phenomenon that the hormonal suites are, generally speaking, not singleton sets¹.

We may wonder whether such complexity is more or less accidental: an outcome of evolution where historical contingencies have created so to speak a series of patches upon bungs upon repairs, resulting in a structure that is much more complicated than it strictly needs to be, “ideally speaking.” Indeed, we might well suppose that an equally effective endocrine system could look like the one depicted in Fig. 2. Here, each target process has its own dedicated hormone, i.e. the system is non-pleiotropic (also, when the concentration of the hormone is low, the process is deemed to go in one direction, and this is reversed when the concentration is high). Whereas the student of endocrinology would be grateful for a highly simplified configuration such as the one shown in Fig. 2, it is not clear that this simplicity would be evolutionarily favoured.

For instance, we observe that a larger number of distinct hormonal messengers is required in the system depicted in Fig. 2. Each hormone needs the support of multiple genes: not only for the production of the hormone itself (a distinction can be drawn here between proteinaceous and non-proteinaceous hormones) but also for its receptor and the signalling

cascades associated with that receptor¹. In principle, the need for additional genes may depress the reproduction rate and this would adversely affect fitness. Thus natural selection would favour pleiotropy. What counts against this argument is that “genomic space” does not seem to be at a particular premium. Even for the smallest mammals (and indeed, for most multicellular organisms), this load does not seem to be a major bottleneck: we need only reflect on the existence of long stretches of apparently non-coding DNA that lie between the genes².

Figure 1 on top of the first page following the title page;

Figure 2 on top of the next page

There are other factors that could account for the persuasive presence of pleiotropy in endocrine systems. For one thing, sheer complexity might constitute a benefit in its own right. According to this line of thinking, networks with redundant connections are more robust, that is, better equipped to withstand perturbations or partial destruction³.

Yet another explanation is that the pattern itself of the pleiotropy of any given hormone could have a particular functional or physiological significance¹. For instance, insulin stimulates processes that collectively act to lower the blood glucose concentration, whereas glucagon has the opposite effects, and cortisol safeguards the availability of energy even under conditions of prolonged restriction of nutrient supplies, at the possible cost of catabolic degradation of all reserves, including the muscle mass (see Fig. 1). Each of these suites can thus be viewed as a coherent set of concerted endocrinological interventions. The interpretation of pleiotropic suites as “physiological modes,” e.g. “metabolic modes” in the example of Fig. 1, is entirely compatible with the idea of complexity as robustness — the difference being that on the complexity explanation per se, the network configuration is more or less arbitrary, whereas the explanation based on modes attributes a particular significance to the system in its ultimate shape.

The notion of endocrine suites as functional modes raises several deep biological questions. One is whether the functional significance that we attribute to these suites, in a more or less *ad hoc* fashion, by inspection and intuition, can be endowed with a rigorous footing. This would endow these modes with some kind of objective reality, even if they are entangled by the basic fact that we noted above: hormones have multiple targets and each target is typically subject to several hormones¹.

Furthermore, if the “suites as modes” concept is tenable, then these modes must necessarily acquire meaning at the whole-organism level. That is to say, the concept presupposes that there is such a thing as unity of purpose at this level, an *organismal harmony*. Hormones are by their very nature well placed to play a central role in maintaining organismal harmony, since they are dissolved in the fluids that bathe the cells and thus can provide a consistent signal at a global level, spanning the entire organism. To view this point in the proper light, we should remember that even multicellular organisms that are barely visible to the naked eye are in fact enormous relative to the basic scale of life, which is that of macro-molecular machinery and cells.

Umwelt: Kennwelt and Wirkwelt

The unitary nature of the organism expresses itself precisely through the fact that certain choices are made at this level, such as fight *versus* flight (short time scales), hunt / forage *versus* rest / digest (intermediate time scales), and somatic growth *versus* reproduction (longer time scales)⁴. When we say that an organism ‘decides’ between the ‘available options’ (i.e. *chooses*), we merely mean that a path is picked from among several alternatives, without necessarily implying any grounding in conscious deliberation, or whatever else it is that humans do to make up their minds (calculations, simulations...). For *choice* in a minimalistic sense, all that is really required is some sort of centralised process that integrates the available information or stimuli that bear on these decisions (Fig. 3). The entirety of options available to an individual organism defines a universe of covert and overt actions which was termed the *Wirkwelt* by Jakob von Uexküll⁵. This *Wirkwelt*, which might be rendered in English as “universe of effect” naturally varies from individual to individual, may also depend on its life stage, and will most certainly be highly specific to the species.

Given a *Wirkwelt*, one can deduce, in principle, which information will be most relevant to making adaptive (that is, fitness-enhancing) choices. There is obviously little to be gained in acquiring and processing stimuli that would not inform a choice among the elements of the *Wirkwelt* (although they could be highly salient in the context of the *Wirkwelt* of a different species). On the other hand, even if we identify stimuli and phenomena that, objectively viewed, ought to be highly salient to the given *Wirkwelt*, we might find they are not sensed by the organism; after all, there is a cost accrued by creating and maintaining, e.g. the sensory organs and the neural circuitry to process the information. This being the case, we would expect natural selection to act so as to find a suitable equilibrium, a compromise between the gathering and processing of “theoretically significant” stimuli and the burden associated with growing the necessary apparatus.

In Uexküll’s natural philosophy, this compromise is known as the organism’s *Kennwelt*⁵. Although it is impossible for a human to know “what it is like” to be a bat, it is eminently possible to study a bat’s *Kennwelt*, its “sensory universe.” Taken together, *Kennwelt* and *Wirkwelt* are known as the organism’s *Umwelt*⁵. Although we have been discussing these ideas with a tacit reference to organisms that have some form of central nervous system (since we began with the example of mammalian endocrinology in mind), the concept of *Umwelt* = *Kennwelt* + *Wirkwelt* extends to all living beings.

Hormones as the mediators of organismal harmony

The argument from organismal harmony explains pleiotropy as a byproduct of unity of purpose at the level of the whole organism, and of hormones acting in concert for these purposes. This notion frequently crops up in informal conversations among (neuro)endocrinologists in common rooms and seminars, where talk of ‘purpose’ is less likely to face censure. In more formal settings, the life scientists would do well to hesitate to invoke purpose, goal, etc., since these terms bear a suspicious teleological burden. We shall revisit the role of teleology in regulatory biology in the closing section. For now, let us just tip our hats to the usual justification: life history evolution provides the underpinnings that make such talk respectable; the informality just spares us the trouble of cumbersome circumlocution.

Figure 3

Organismal harmony encompasses the coherence of the organism's *Umwelt*, which requires concerted coordination of activity across different organs and tissues. Local adjustments are made to accord with the global (i.e., *Umwelt*-level) decision. The latter is often analysed in terms of a *trade-off*: the fitness that alternative **A** might have contributed is bargained against the fitness associated with alternative **B**^{4,6}. It is *prima facie* reasonable to suppose that the trade-offs achieved are balancing points at which fitness is maximised; however, the application of optimisation ideas to evolutionary scenarios can be fraught⁷.

The term *life history trade-off* is used primarily when discussing the intermediate-to-lifetime scales, although the trade-off concept is relevant across the entire continuum of time scales. For instance, the pond snail *Lymnaea stagnalis* maintains a particular concentration of sodium ions in its h emolymph; under conditions of low sodium concentrations in the external medium, such maintenance requires the uptake of external sodium ions by active mechanisms in the integument to make up for passive and urinary losses⁸. The substantial fitness benefits of maintaining a stable electrolyte environment in the fluid that bathes the tissues are traded off against the opportunity costs incurred when energy is invested in transporting ions against the overall ambient/h emolymph gradient and molecular building blocks are invested in synthesising the molecular machinery that mediates this transport.

As the h emolymph sodium ion concentration falls, the snail's neuro-endocrine system secretes Sodium-Influx Stimulating peptide⁸. The resulting elevated levels of this peptide hormone in the h emolymph represent the *drive* to maintain a (nearly) constant h emolymph concentration. We might say that the hormone effectively *urges* the integument to take up sodium.

Indeed, the word *hormone* derives from ὁρμῶν (hormon), the present participle of ὁρμᾶν (horman), 'to set in motion' or 'urge' which in turn goes back to ὁρμή (horme) which might be translated as 'onset,' 'urge,' or 'impulse.' In the teachings of the Stoic philosophers, this ὁρμή could be aroused by an external stimulus or internal stirrings of the body^{9,10}; it is perhaps not too fanciful to regard these arousals as the Stoic counterpart of Uexk ull's *Kennwelt*, the sum total of sensory stimuli available to a given organism (Fig. 3).

The way the Stoics rendered ὁρμή in Latin tells us much about their thinking: *appetite* for instance, deriving from *adp t tio*, 'that which is sought or desired;' and similar terms that ultimately gave rise to our *impulse* and *impetus*⁹. The Stoics regarded these impulses as natural processes which happen in and around our bodies — quite independently of our will. The distinction between what is and what is not under the control of human volition played a central role in Stoic philosophy⁹. Accordingly, Stoics praised above all the ability to control these natural urges by exerting judgement and self-discipline, which lend or withhold assent (συγκατάθεσις = synkatatheste) to these drives¹⁰.

It may be of some interest that the Stoics did not always differentiate clearly between *afferent* and *efferent* signals, as the ὁρμή was both inciting stimulus arising from physiological or external events and the urge to act in the natural way, following what we might envisage as the "path of least resistance" for the organism^{9,10}. Channeling and regulating this natural flow thus becomes tantamount to asserting our humanity, given that beasts were supposed to simply go with this flow.

Figure 4

While the distinction between afferent (incoming information) and efferent (outgoing information) is key to the central nervous system (CNS), in the endocrine system we find that the bearers of ὁρμή do in fact play a dual role. Factors such as insulin and leptin can be viewed as efferent in the sense of directing peripheral physiology, but they act at the same time as afferent to the CNS, where hypothalamic centres sense their presence and adjust organismal homeostasis in the basis of the information received¹¹.

Modern science defines hormones, the molecules that mediate ὁρμή, more precisely: they are humoral signalling molecules that are distributed throughout the body via the blood or hæmolymph compartment, as opposed to *paracrine*, *autocrine*, and *juxtacrine* factors, which are confined to the immediate tissue environment¹. All such factors are called *first messengers* in contradistinction to *second messengers*, signalling molecules that trigger intracellular signalling cascades. Synaptic communication by means of neurotransmitters can be regarded as a special case of paracrine/autocrine signalling¹. First messengers usually evoke cellular responses by engaging receptors situated at the exterior side of the membrane of the target cell, but they may also engage intracellular receptors, in which case signalling is *intracrine*¹.

Hormones play a central role as regulators of life history trade-offs. For instance, testosterone mediates the trade-off between reproduction and immunity in the bird *Junco hyemalis*⁴; follicle-stimulating hormone (FSH) mediates the trade-off between egg number and egg size in the lizard *Uta stansburiana*¹³; and juvenile hormone (JH) mediates the trade-off between flight capability and reproduction in the insect *Gryllus firmus*⁶. The choice between, on the one hand, investments in somatic mass and endurance, and on the other hand, direct investment in offspring, is among the most consequential life history trade-offs: investment in the former can, with suitable discounting, be leveraged into more of the latter, albeit later in life⁷. It is this juvenile/larva → adult transition (or, more generally, non-reproductive → reproductive phase transitions, possibly iterated) where pleiotropy manifests itself most dramatically, since the different life stages often entail profound physiological, anatomical, and biochemical differences. The Endopterygota (or holometabolic insects) constitute an example *par excellence*, with individuals undergoing a radical metamorphosis through a pupal stage in which the entire body mass is reorganised. The pleiotropic suite of JH in the holometabolic insect *Drosophila* includes numerous targets, involving virtually all aspects of the phenotype, such as metamorphosis, development and differentiation of adult body structures, reproductive organs, pheromone production, locomotor and courtship behaviour, division of labour in social insects, brain structure and function¹⁴. The radical realignment of purpose at the whole-organism level is a change of key of the organismal harmony, and correlates with an extensive suite of JH targets. Perhaps ‘purpose’ here begins to acquire a more respectable sense, imposed by the evolutionary forces in the background.

From an evolutionary point of view, trade-offs represent balancing points between evolutionary pressures⁴. At the level of neuro-endocrine regulation, the trade-off is a balance of opposing drives, negotiated by hormones and other first messengers¹. Combining these two perspectives, we perceive first messengers to be *tokens* of the evolutionary pressures (i.e. fitness differentials) that govern such trade-offs¹. However, if first messengers (and perhaps second messengers as well) are truly tokens of selection, it would seem that we

have quite a bit of explaining to do. For one thing, humeral messengers necessarily act over time scales several orders of magnitude shorter than the typical life span of the organism, whereas the latter act over time scales several orders of magnitude longer. Thus we are led to wonder whether “hormones betoken evolution” is a true biological principle or simply an elegant heuristic to arrange our facts and figures.

Figure 5

Of q -systems and Q -systems

Let us begin by stating that biological systems exhibit regulation and that this regulation is organised in the form of, among other things, feedback loops¹⁵. What do these feedback loops do? They tend to drive a biological system away from a given physiological state with an urgency that is proportional to the “maladaptiveness” of that state, and toward more adaptive states (but for the cost-trade off proviso we noted above). How do hormones act? They do not typically carry the energy needed to drive the responses they elicit (in the sense that the required ΔG would be liberated upon their degradation in the target tissue). Their role is strictly that of a *messenger*, triggering certain responses in cells that carry receptors for the hormone at hand, called the *target cells*. Gene expression may be altered, transporters may be activated, membrane permeabilities changed, biochemical pathways set in motion or slowed down, development of certain tissues (e.g. with a reproductive function) started or arrested, and so on¹. This messenger / target duality is apparent from Fig. 1 where the coloured lines are the messengers and the ‘stops’ are the regulated processes.

If we look at the messenger / target duality from the point of view of physiology, we will be able to identify certain points where the messengers exert their influence on the dynamics of the system (Fig. 4). Let us call these points the *actuators*. These are the particular physiological parameters that change value in response to an endocrine signal, and in this way alter the trajectory of the physiological dynamics. In mathematical terms, one might surmise that the gradient of the dynamics with respect to these actuator parameters contains the necessary information to determine the connectivity structure of the control network. Indeed, it turns out that this insight is a key step toward objectively defining suites in terms of coordinated organismal modes¹⁷.

The classification of variables as either actuator variables or physiological variables ‘proper’ is somewhat arbitrary, but it is elucidated by a familiar problem in mathematical biology. Physiological variables ‘proper’ (the coloured boxes in Fig. 3) are generally straightforward to model; this is not to say that the task is without its challenges, but simply that basic physicochemical principles such as conservation laws tend to go a long way toward mathematical *closure* (a term of art mathematicians use to indicate that the dynamics of the state can be fully specified on terms of the state itself). This closure often proves to be considerably more problematic in the case of the actuator variables. To specify these as functions of time, we require a mathematical model of the neuro-endocrine system (the black boxes in Fig. 4), whose output dictates the values of the actuators.

The simplest way to arrive at such a model begins by identifying the afferent inputs to the system. This would include physiological variables whose values are relevant on the behaviour of the actuator we are trying to model. Usually some sort of sensor mechanism provides the afferent signals. Next, we identify the efferent outputs, which include the actu-

ator variable we are interested in. We then try to find a mathematical function that maps the input to the output.

Figure 6

For example, as shown in Fig. 4, the black box marked ‘ β ’ representing the pancreatic beta cells that monitor glucose and secrete insulin in response¹. If we take the insulin secretion rate to represent the actuator variable in our model, we could propose a simple function $\mathbb{R} \rightarrow \mathbb{R}$ that maps glucose concentration to the insulin secretion rate¹⁸. Insulin affects the rates of a number of physiological processes, as shown in Fig. 1, and accordingly we need to keep track of the dynamics of insulin in the blood and the manner in which these more distal actuators depend on the blood plasma insulin concentration¹⁸.

The black box marked ‘CNS’ (central nervous system) in Fig. 4 poses a greater challenge. At the neuroanatomical level, we can locate the structures in the brain that mediate this rather more intricate input-output mapping¹²; some of these are shown in Fig. 5. Both the inputs and outputs of this system have been identified¹¹, allowing us to draw up a putative wiring diagram (Fig. 6).

In general, if we have n inputs and m outputs, our first attempt would be to construct a function $\mathbb{R}^n \rightarrow \mathbb{R}^m$ on the basis of the available data. Let us consider the case where we find this to be impossible (otherwise there is not much to say). Then if the system is *finitely deterministic* (which is a reasonable assumption for a regulatory system) there is a finite integer q such that there *does* exist a function $\mathbb{R}^{n+q} \rightarrow \mathbb{R}^m$ that describes the m actuators (this is the system’s output map¹⁹). Moreover, the coordinates in the additional q dimensions of the domain of this mapping cannot all have the same values, at all points in time where we measured; for otherwise we could set $q = 0$ which would imply that the $\mathbb{R}^n \rightarrow \mathbb{R}^m$ map should have worked after all. Incidentally, time resolution plays a central role in this argument, in the sense that coarse-graining the resolution in time (i.e. averaging over smaller time windows) will allow us to reduce q , perhaps even all the way to zero. However, in general we have to be prepared to deal with the case $q > 0$; importantly, we will have to account for the time-varying behaviour of these coordinates.

The term ‘coordinates’ is somewhat misleading, in the following sense: given one example of the desired map $\mathbb{R}^{n+q} \rightarrow \mathbb{R}^m$, we can precompose this map with a map that is arbitrary as long as it is invertible and the identity in the first n arguments (but not the final m), and we have constructed *another* $\mathbb{R}^{n+q} \rightarrow \mathbb{R}^m$ map that serves the purpose just as well. In other words: we have to think of whatever exists in these additional q dimensions as a single, coordinate-independent, entity. The proper name for this entity is the *state*. The rules that describe how the state evolves over time are collectively called the state’s *dynamics*¹⁹. As a special case, these rules might actually allow the state to remain the unchanged as time elapses. In this special, so-called *static*, case we are thrown back to the much simpler $\mathbb{R}^n \rightarrow \mathbb{R}^m$ mapping.

While there do exist general abstract methods to identify the state¹⁹, they typically require many more observational data points than are available, and for this reason mathematical modellers have to resort to heuristics. One of these is to identify key physiological variables, each of which can be deemed likely to fulfil the role of a “ q -coordinate” (i.e., a state variable). Obviously this requires judgement based on additional knowledge of the sys-

tem (and biology in general)²⁰. The dynamics of the state are then specified, for instance in the form of differential equations, which is a formalism that usually suits the causal structure of the subject²⁰. The effectiveness of this approach is such that it has become virtually synonymous with mathematical modelling²⁰.

In the case at hand, the problem is not really that we struggle to identify candidate state variables. On the contrary, the problem is rather an embarrassment of riches. We know quite well how to describe action potentials, synaptic communication, hormone secretion, receptor binding and signalling cascades, and so on²⁰. However, if we were to incorporate all this knowledge into a mathematical model, we would quickly find that q has a disconcerting tendency to grows beyond any reasonable bound.

Well, who is to say where we should put a “reasonable bound”? There is a school of thought that holds that we should not worry when q runs into the thousands (or even tens or hundreds of thousands...), as modern computing technology has eliminated the problem. The computer can simulate it all, and if the computer is big enough, it will even do so faster than real time. This attitude of “simulate the lot” is indubitably a defensible philosophical stance. In practice, however, there are several snags. Assigning reliable values to all parameters is one of them (but perhaps this becomes *less* important when q is very large). Another snag is that, besides the mysteries of the original biological system, the modellers have now created an equally impenetrable *in silico* system, an entity that lives in the computer which may or may not be a faithful simulacrum (we hope so, but even then we never know if it is accurate for the “right” reasons). If we are thus driven to formulate a tractable (which would also mean: a lower-dimensional) theory to account for the simulation model — a *meta-model* — it will dawn on us that we cannot escape a fundamental question: what is the smallest q that allows a map $\mathbb{R}^{n+q} \rightarrow \mathbb{R}^m$ to perform with sufficient accuracy? Perhaps we ought to have started with *that* question.

The latter “minimal state” stance distinguishes itself from the “just simulate the lot” approach in several ways. We anticipate that, should the min- q problem have a solution, it will also be “the” solution in the sense that any other solutions we might find will only differ from the original one in more or less unimportant ways. Moreover, if we are able to define a coordinate system for the minimal q -space in such a way that the individual coordinates permit biological interpretation, then so much the better, but we are not necessarily expecting or demanding this. In other words, the minimal state variables will probably not correspond to familiar quantities such as concentrations of solutes in certain compartments, receptor densities, membrane potentials, and so on. The truth is that such a guarantee never inhered in the state concept, notwithstanding common prejudice which, as we have pointed out, merely springs from the popularity of one particular modelling heuristic.

There are various approaches to the min- q problem. One is applicable when adepts of the “simulate the lot” school have in fact gone ahead and produced a dynamical model with an ultra-high-dimensional state. Let $\mathbf{X} \in \mathbb{R}^Q$ denote thus model’s state (where we have written the dimension as a capital Q to remind us that $Q \gg 1$). From simulations of this model, we may discover that the state is mostly confined to a low-dimensional subspace, in the sense that we can write $\mathbf{X} = \mathbf{\Xi}(\mathbf{x}) + \mathbf{c}(\mathbf{X})$ where $\mathbf{x} \in \mathbb{R}^q$ with $q \ll Q$ and $\|\mathbf{c}(\mathbf{X})\| \ll \|\mathbf{\Xi}(\mathbf{x})\|$. For each $\mathbf{X} \in \mathbb{R}^Q$ there will be an $\mathbf{x} \in \mathbb{R}^q$ such that $\|\mathbf{c}(\mathbf{X})\|$ is minimised; this defines a function $\mathbf{P}(\mathbf{X}) \in \mathbb{R}^q$ (if the minimiser is not unique we will have

to break the tie somehow). Let us suppose that the original Q -system dynamics can be written in the following form:

$$\frac{d}{dt}\mathbf{X} = \mathbf{F}(\mathbf{X}, \mathbf{u})$$

where $\mathbf{u} \in \mathbb{R}^n$ is the vector of monitored variables. The function \mathbf{F} is what the modelling team have provided us with. Then a proposal for the reduced dynamics of a putative “ q -system” would read:

$$\frac{d}{dt}\mathbf{x} = \mathbf{P}_{\mathbf{X}}(\Xi(\mathbf{x})) \cdot \mathbf{F}(\Xi(\mathbf{x}), \mathbf{u}),$$

where $\mathbf{P}_{\mathbf{X}}$ denotes the derivative of \mathbf{P} . The point of this expression is mathematical closure: the right-hand side *only* depends on \mathbf{x} (and \mathbf{u} , of course, but *not* on \mathbf{X}). The dynamics has been reduced to a q -dimensional arena.

A solution $\xi(t) \in \mathbb{R}^q$ satisfying the reduced dynamics $\frac{d}{dt}\mathbf{x}$ would induce an *approximate* solution of the original system $\mathbf{X}(t) \approx \Xi(\xi(t))$. There is no guarantee that this induced *Ersatz* solution $\Xi(\xi(t))$ will accurately mimic the behaviour of the Q -system. The *Ersatz* solution is exact only in the exceptional case where the Q -system’s trajectory was such that $\|\mathbf{c}(\mathbf{X}(t))\| = 0$ for all t . Barring this non-generic case, we have to expect that $\Xi(\xi(t))$ may deviate from $\mathbf{X}(t)$ in important qualitative ways, e.g. with different stability or recurrence properties. Moreover, from a physics point of view, the *Ersatz* may violate key principles such as conservation properties (even when the Q -system did not). The latter type of difficulty can be resolved by suitably constraining the construction of either the map $\Xi(\cdot)$ or $\mathbf{P}(\cdot)$; the details, which are usually problem-specific and fairly technical, are beyond the scope of the present paper.

In any case, whether or not the reduction to a q -system has been successful must be judged on whether the output is described with sufficient accuracy. In the Q -system this would be represented as a $\mathbb{R}^{n+Q} \rightarrow \mathbb{R}^m$ output map $\mathbf{G}(\mathbf{u}, \mathbf{X})$. The reduced approximation would then be $\mathbf{G}(\mathbf{u}, \Xi(\xi(t)))$ and it is really the fidelity of this output that should be used to assess the success of the state reduction. The special *static* case is the one where $q = 0$; for then we are back at $\mathbf{G}(\mathbf{u}, \Xi(\xi(t))) \equiv \mathbf{G}(\mathbf{u}, \bar{\Xi}) \equiv \mathbf{G}_{\bar{\Xi}}(\mathbf{u})$ as an $\mathbb{R}^n \rightarrow \mathbb{R}^m$ mapping.

Hormones as tokens of evolutionary pressures

Let us suppose that a Q -to- q reduction as outlined above is not only feasible but also that any reasonably comprehensive Q -system will reduce to a q -system that differs from the reductions of any other such Q -system only in a non-essential manner (at least relative to a given output space of actuator variables). We know little about the form the resulting q -systems might take. The foregoing arguments have prepared us to assume that the state variables of such a system need not, and in all likelihood will not, admit a direct biological interpretation in terms of the components that were explicitly represented in the Q -system. Instead, for the q -system we anticipate that we will have variables (a.k.a. phase coordinates or dynamic degrees of freedom) of a more abstract nature. To see what form these might take, it is perhaps fruitful to return to the concept of hormones as mediators of life-history trade-offs, and, more generally, of evolutionary pressures⁴.

An clear example where the physiology appears to resonate perfectly with the concept of hormones as tokens of selective pressure is the regulation of the blood plasma glucose titre. Both low and high “blood sugar” lead to serious adverse effects²¹, some of which are listed in Fig. 7. These morbidities will have a negative impact on the lifetime reproductive output of the organism, and hence evolution is expected to favour mechanisms that will induce a tendency to avoid them. These are known as *homeostatic mechanisms*¹⁵.

Figure 7

Figure 8

The evolutionary “landscape” associated with blood glucose homeostasis is sketched in Fig. 8; the optimum value is shown as the lowest point in a valley. The physiological state is pictured as a ball that runs down the slopes of the values. The driving forces (gradients of the landscape) are depicted as arrows corresponding to the hormones that betoken them.

We can take the mechanical analogy one step further and write down the corresponding classical mechanics for this problem, in order to obtain specific equations for a tentative q -system. (For good measure, we would first submerge the system in viscous oil to avoid oscillations and spare ourselves an extra differentiation step; also, what is depicted is actually an inverted fitness landscape, with the highest fitness value occurring at the lowest point.) Besides being a picturesque metaphor, does this sketch offer any scope for a fully-fledged theory of endocrinology at the “ q -system level”? Let us examine a few of the more serious objections.

First, the insulin-glucagon system might just be a felicitous but otherwise atypical example that was cherry-picked as it happens to accord well with the theoretical ideas that are being put forward here. However, the approach is applicable to various physiological systems, including hydro-mineral regulation, body temperature control, and regulation of macro-chemical body composition^{17,16}.

Furthermore, it is not clear what is actually set out along the abscissa in Figs. 7 and 8. Ostensibly it is the plasma glucose concentration, but this quantity fluctuates during the lifetime of the organism, whereas lifetime reproductive output is a fixed number, for a given life history. We could fix this by taking the abscissa to indicate the blood glucose level *averaged* over the life time. However, the following approach is a bit neater, and also more suited to later generalisations. Let $x(t)$ denote the glucose level and $L(\cdot)$ a U-shaped function as represented by the “valley” in Fig. 7. Then consider the quantity

$$U = \int_{t_{\text{birth}}}^{t_{\text{death}}} L(x(\tau))d\tau.$$

Fitness marginalised to x (i.e. averaging out all traits except the above integral) is then a monotonically *decreasing* function of U (which we might term the *unfitness*).

This U -formalism can be generalised in the obvious manner to physiological states of any dimension, with the function L depending on all of the elements of the corresponding state vector. Life-stage dependent variation of the fitness landscape is accommodated in this extension, since life stage is subsumed under the concept of state, as is, for that matter, time dependence in general¹⁹. Although the generalisation is entirely straightforward from a mathematical point of view, one may object, from a biological perspective, that L is given,

as it were, too much work to do. We are led to suppose that L encodes the information regarding the connection between physiology and fitness, which prompts us to ask whether L is observable in principle. It is; but the observational data would have to be time-resolved several orders of magnitude *below* the typical intergeneration time T_g but also span a period several orders of magnitude *longer* than T_g , and extend over a *vast* ensemble of distinct life histories¹⁷. While one may legitimately object that the theory relies on our ability to overcome this huge practical obstacle (perhaps eventually; for microorganisms the acquisition of such a data set may become a reality in the near future), the point is rather that the same fundamental difficulty must be overcome by any sufficiently detailed theory relating evolution to physiology *sensu lato* (i.e., physiology in the widest sense of the word).

Another objection relates to the way in which the gradients in Fig. 8 were identified with actuators (and labelled with the names of hormones that form the bridge between the actuators and the gradients of L). In this case, it suffices to inspect Fig. 1 (or any other chart or table conveying the same information) and identify actuators that would tend to raise blood glucose or lower it, respectively. Strong depressing actuators are “stations” along the “insulin line” whereas strong elevating actuators lie along the “glucagon line” — there is nothing coincidental or surprising about this, but the question we are really interested here is how we could have worked this out, even if we did not remember these facts from our endocrinology book. The objection, then, is that there ought to be a systematic way to determine the coupling strengths that should prevail, intervening between the gradients of L and the “pushes” or “pulls” on the actuators.

This is not a trivial problem since a given actuator will affect a certain subset of the physiological state variables in the positive direction, and another subset in the negative direction (one of these subsets may be empty). Conversely, a given physiological state variable may be governed by more than one actuator (it has to be at least one, lest the physiological variable be insusceptible to regulation). As a result, a deadlock may arise between the simultaneous pushes and pulls exerted on a given actuator by several physiological state variables. We may well surmise that some sort of compromise is reached for each individual actuator — but then a given physiological state variable could end up being governed by actuators that are subject to such compromises and therefore not necessarily steered in the “correct” direction, at least as far as regulation of the physiological state variable at hand is concerned. This problem can be solved; we shall not delve into the particulars here¹⁷.

Intrinsic and extrinsic teleology

The conclusion we reached in the previous section hinged on the supposition that, loosely speaking, the function L adequately encodes physiological goals. However, this wording seems presuppose an inherent goal-orientedness in the regulatory systems of living beings. Purposefulness, or at any rate the appearance of it as witnessed by our readiness to use the language of purposefulness in describing living things, has always been a contentious issue^{22,23,24}.

As we saw, L is essentially an inverted version of the fitness landscape. The link between L and fitness (via the unfitness integral U) is an explicit representation of how the action of natural selection accounts for the emergence of apparent goal-orientation in biological organisms. The fitness landscape may be reified as the result of the myriad interactions

between the organism's realised life trajectory and its environment. To specify it in detail would certainly be a tall order encompassing as it does all of a given species' biology and ecology. Yet this is only a matter of practical feasibility; the fundamental point is that the interactions that make up the fitness landscape are of a strictly physico-chemical nature. In this sense, the fitness landscape "simply is there" as an objective reality. If we accept our function L as a reflection of this fitness landscape, we will have reified it and thus dispelled the spectre of teleology.

As an example of the language of purposefulness, the local minima of the function L can be referred to as "set points," or, even more outrageously, "desired" or "optimal" values. If we have already conceded that the existence of natural selection reifies L , then there is no real conceptual problem: we are simply dealing with a mathematical function and we may attach to its minima any labels that tickle our fancy. On the other hand, we do have a problem if we view *set point* as the primitive concept: e.g. as the end-point to which the system is continually striving. In an engineered artefact such as a throttle governor, we know who it was that set the set point: either the engineer who designed it, or perhaps the end user of the instrument. But regulating loops in living systems are not engineered artefact. Or are they?

Whenever we discern such apparent purposefulness in biological systems, we could claim this is an expression of $\tau\acute{\epsilon}\lambda\omicron\varsigma$ (telos), a final end point or inherent aim which is already intrinsic in the nature of the thing. Classical antiquity adhered to a doctrine of *natural teleology* which endowed the laws of nature itself with such intrinsic purpose²⁵. This is perhaps best viewed as a natural, unforced unfolding rather than a conscious striving. Intrinsic $\tau\acute{\epsilon}\lambda\omicron\varsigma$ is cut of the same cloth as the Stoics' *impetus* or $\acute{o}\rho\mu\eta$: the same kind of natural philosophy is at work here, that ignores the microscopic nature of the universe and explains phenomena as either unforced natural unfolding, or else forced alterations thereof, where the forces are imposed by beings that can exert volition.

Whilst $\tau\acute{\epsilon}\lambda\omicron\varsigma$ represents an intrinsic or inherent kind of purposefulness, there is also the extrinsic kind that resides in the use and design which beings capable of foresight impose on objects or fellow creatures. The distinction is dulled in Western culture, inasmuch as intrinsic and extrinsic purposes have a tendency to come to much the same thing in the context of monotheistic faith, such as the Abrahamic religions, which presuppose a personal, omnipotent, and sometimes benevolent supreme being. (Perhaps this is also true for other belief systems, insofar as power or ownership, volition, and personal consciousness are ascribed to super-natural entities.)

As the intrinsic and extrinsic purposefulness become conflated, one risks losing sight of the fundamental distinction between the *apparent* design of a biological control loop and the deliberate *engineering* design of e.g. a throttle governor. Both come to be seen as an expression of a conscious mind that has arranged apparatus (biological or brass) based on a kind of reasoning backwards from desired final states. The appeal of "Intelligent Design" touted as a creationist (pseudo-)science reminds us that, in the context of a monotheistic faith, the appearance of intrinsic purposefulness automatically entails the further presence of extrinsic purposefulness.

The muddling of intrinsic and extrinsic purposefulness stands in stark contrast to the doctrine of *natural teleology*, which maintains a strict distinction. This does not imply that natural teleology is necessarily atheistic or agnostic. Natural teleology can be seen to be compatible with deism or suitably adapted forms of theism, such as pantheism, or even

monotheism, provided that the divinity is primarily conceived of as the first efficient cause. The rejection of natural teleology should therefore not be taken to be a disparagement of any side in religious debates.

The U integral in the previous section calls to mind the variational principles that play a prominent role in contemporary theoretical physics²⁶. It is tempting to infer from the notion of global optimisation that some kind of teleology could serve as a central organising principle of all natural sciences; Max Planck for one believed that this might be the case²⁶. Nevertheless, there is no reason to suppose that mathematical formalism is a suitable well-spring of metaphysical truth.

The doctrine of natural teleology is still occasionally put forward as an explanation for apparent purposefulness in the life sciences²⁴. It is difficult to decide what to make of this: we cannot possibly presume to know how a Stoic would respond when confronted with our current understanding of molecular genetics, cell biology, and development. There is clearly an informational structure that resides in a germ (e.g. a zygote) that unfolds (“to develop” means to unfold or unwrap) to express its inherent nature. All of this would be in accordance with the intrinsic nature of $\tau\acute{\epsilon}\lambda\omicron\varsigma$ as the Ancient understands it. The question is whether the interpretation of DNA as $\tau\acute{\epsilon}\lambda\omicron\varsigma$ would suffice for the classic teleologist: are the interactions between the molecules, such as modern-day chemists and physicists describe these, enough, or would there still be the presumption of an underlying $\tau\acute{\epsilon}\lambda\omicron\varsigma$ that drives it all forward (slightly reminiscent of the *vis vitalis* — ‘life force’ — of the intervening ages)? In the final analysis, we do not yet have a sufficiently complete picture of what makes the apparent laws of nature the way they are, and thus, strictly speaking, we are not yet in a position to put teleology to rest once and for all, although we are tempted to agree with Laplace and say *nous n’avons pas besoin de cette hypothèse-là*.

References

- [1] B. Kleine and W. G. Rossmanith (2016) *Hormones and the Endocrine System*. Springer, Berlin.
- [2] J. A. Shapiro (2017) Biological action in Read–Write genome evolution. *Interface Focus* **7**: 20160115; DOI: 10.1098/rsfs.2016.0115.
- [3] J. M. Carlson and J. Doyle (2002) Complexity and robustness. *PNAS* **99**: 2538–2545.
- [4] E. D. Ketterson and J. W. Atwell (2016) *Snowbird: Integrative Biology and Evolutionary Diversity in the Junco*. University of Chicago Press, Chicago.
- [5] Jakob von Uexküll (1934) *Streifzüge durch die Umwelten von Tieren und Menschen*. Rowohlt, Hamburg.
- [6] A. J. Zera and L. G. Harshman (2001) The physiology of life history trade-offs in animals. *Ann. Rev. Ecol. Syst.* **32**: 95–126.
- [7] H. A. van den Berg (2015) *Evolutionary Dynamics: The Mathematics of Genes and Traits*. Institute of Physics, London.
- [8] N. D. de With and H. A. van den Berg (1992) Neuroendocrine control of hydromineral metabolism in molluscs, with special emphasis on the pulmonate freshwater snail, *Lymnaea stagnalis*. *Adv. Comp. Endocrinol.* **1**: 83–99.
- [9] J. Rist (1977) *Stoic Philosophy*. Cambridge University Press, Cambridge.
- [10] J. A. Stevens (1996) Impulse and animal action in stoic psychology. *News. Soc. Ancient Greek. Phil.* **204**: 12–1996.

- [11] A. K. Hewson, L. Y. C. Tung, D. W. Connell, L. Tookman, and S. L. Dickson (2002) The rat arcuate nucleus integrates peripheral signals provided by leptin, insulin, and a ghrelin mimetic. *Diabetes* **51**: 3412–3419.
- [12] M. W. Schwartz, S. C. Woods, D. Porte Jr, R. J. Seeley, and D. G. Baskin (2000) Central nervous system control of food intake. *Nature* **404**: 661–671.
- [13] B. Sinervo (1999) Mechanistic analysis of natural selection and a refinement of Lack’s and Williams’s principles. *Am. Nat.* **154**:S26– S42.
- [14] T. Flatt, M.-P. Tu, and M. Tatar (2005) Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. *BioEssays* **27**: 999–1010.
- [15] M. Zimmermann (1989) General principles of regulation. In: R. F. Schmidt and G. Thews, editors, *Human Physiology*, chapter 15, pages 324–332. Springer, Berlin.
- [16] H.A. van den Berg, Yu. N. Kiselev, and M. V. Orlov (2015) Homeostatic regulation in physiological systems: A versatile Ansatz. *Math. Biosci.* **268**: 92–101.
- [17] H. A. van den Berg (2019) *Hormones as Tokens of Selection: Qualitative Dynamics of Homeostasis and Regulation in Organismal Biology*. CRC Press, Boca Raton.
- [18] Y.-F. Wang, M. Khan, and H. A. van den Berg (2012) Interaction of fast and slow dynamics in endocrine control systems with an application to β -cell dynamics. *Math. Biosci.* **235**: 8–18.
- [19] L. Padulo and M. A. Arbib (1974) *System Theory*. Saunders, Philadelphia.
- [20] H. A. van den Berg (2011) *Mathematical Models of Biological Systems*. Oxford University Press, Oxford.
- [21] R. I. G. Holt and N. A. Hanley (2012) *Essential Endocrinology and Diabetes*. Wiley-Blackwell, Hoboken.
- [22] M. Grene and D. Depew (2008) *The Philosophy of Biology: An Episodic History*. Cambridge University Press, Cambridge.
- [23] J. Jacobs (1986) Teleology and reduction in biology. *Biol. Philos.* **1**: 389–399.
- [24] T. Nagel (2012) *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. Oxford University Press, Oxford.
- [25] C. Allen (2009) Teleological notions in biology In: *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2009/entries/teleology-biology/>.
- [26] W. Yourgrau and S. Mandelstam (1979) *Variational Principles in Dynamics and Quantum Theory*. Pitman, London.

Figure captions

Figure 1. Eight key hormones involved in mammalian energy metabolism, whose suites are portrayed here as lines in a subway system: stops along each line correspond to hormonal targets.

Figure 2. Alternative imaginary (much simplified) endocrine infrastructure for the system shown in the previous figure. (The colours do not necessarily correspond with those in the

Figure 3. The *Umwelt* concept. The organism's sensory modalities together make up the *Kennwelt*, The *Wirkwelt* encompasses the organism's decisions on its behaviour (both overt and covert), development, physiology, and metabolism. The informational cascade of sensory integration, processing, and selection defines an organismal unit and thus underlies organismal harmony. The flow into the central processing machinery is referred to as *afferent*, that exiting this machinery as *efferent*.

Figure 4. A subset of the endocrine factors and targets from Fig. 1, showing in more detail the processes governed by these factors. Dotted lines show regulatory actions; dashed lines show information flows toward the central nervous system (CNS); and heavy drawn lines show fluxes and biochemical transformations. The latter are the "actuators" of the physiological dynamics. The regulatory component is shown as black boxes, whereas the physiology (*s.l.*) they control is shown as coloured boxes. The boxes labelled α and β represent pancreatic alpha and beta cells.

Figure 5. Regions in the rat brain mediating the integration of energy homeostasis, arousal, and appetite (feeding drive). The sagittal cross-section (top) shows where the two transversal sections (bottom) were cut (rostral means toward the front of the animal). The left transversal section exhibits the "input" regions: ARC, arcuate nucleus; DMN, dorsomedial nucleus; VMN, ventromedial nucleus. The right transversal section shows the "output" regions: PVN, paraventricular nucleus; LHA, lateral hypothalamus; PFA, perifornical area. Also shown are: TH, thalamus; FX, fornix; OC, optic chiasma.

Figure 6. The neuro-endocrine interface pathways associated with the areas depicted in the previous figure. The arcuate nucleus receives several humoral inputs, among which are the levels of the hormones insulin, leptin, and ghrelin, as well as the key nutrient glucose. This nucleus contains two populations of neurones with opposing actions on feeding drive, energy use, and energy preservation, the NPY/Agrp cells and the POMC/Cart cells. These project to various other areas as shown and modulate the release of the factors indicated. The outputs are both neural pathways to elsewhere in the brain, influencing behaviour, and endocrine outputs via the hypothalamic-pituitary axis.

Figure 7. Adverse physiological effects accompanying hypoglycaemia (low blood sugar) and hyperglycaemia (high blood sugar).

Figure 8. Ball-rolling-down-a-hill analogy for selective pressures.

Figures

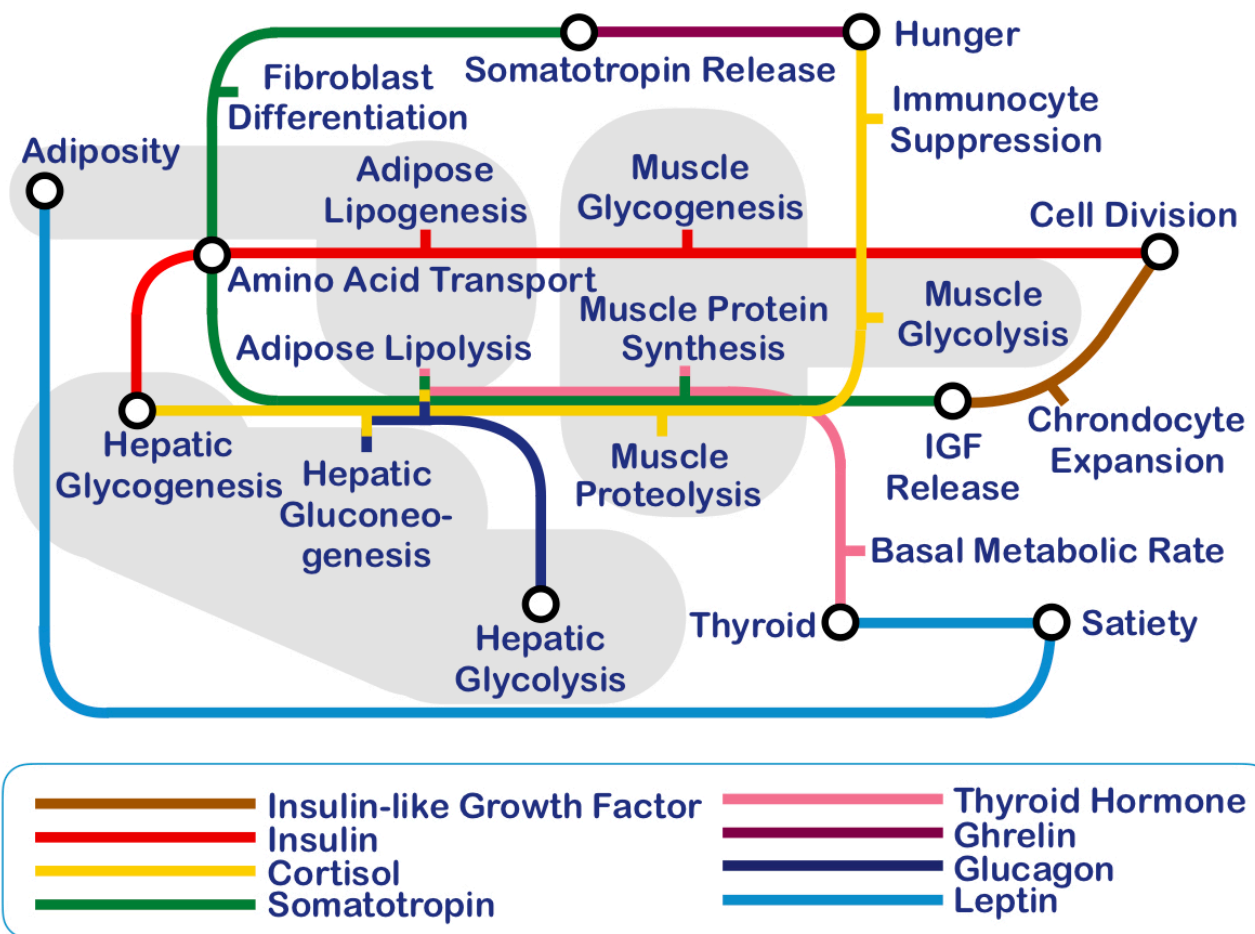


Figure 1

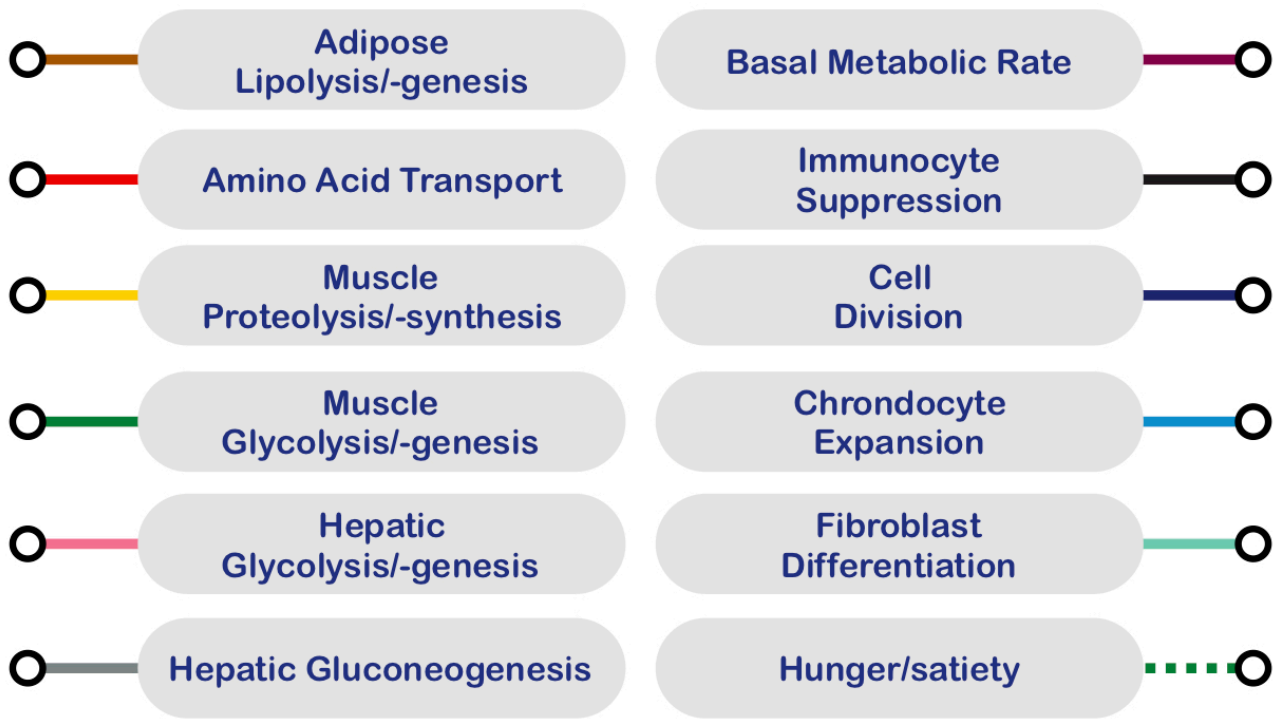
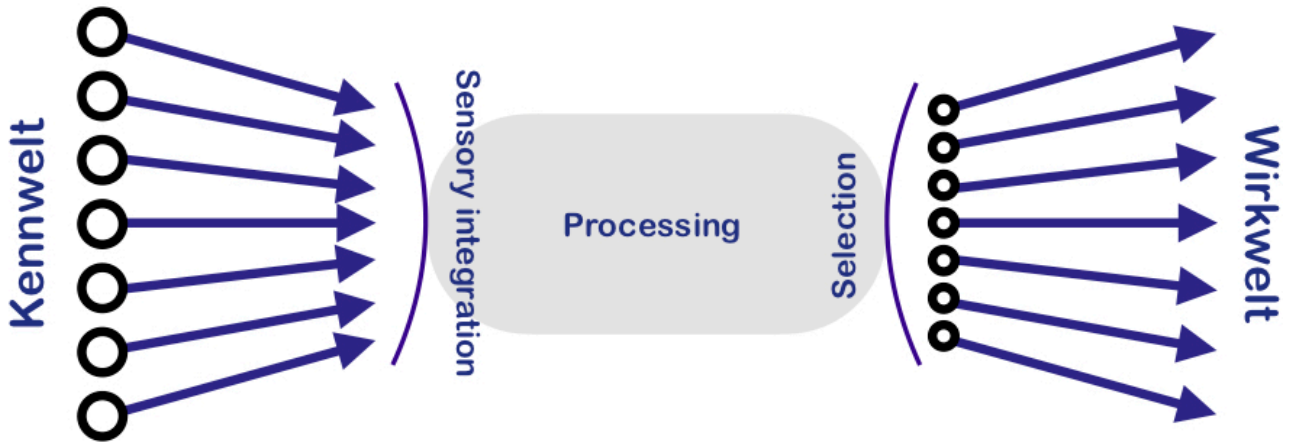


Figure 2

Figure 3



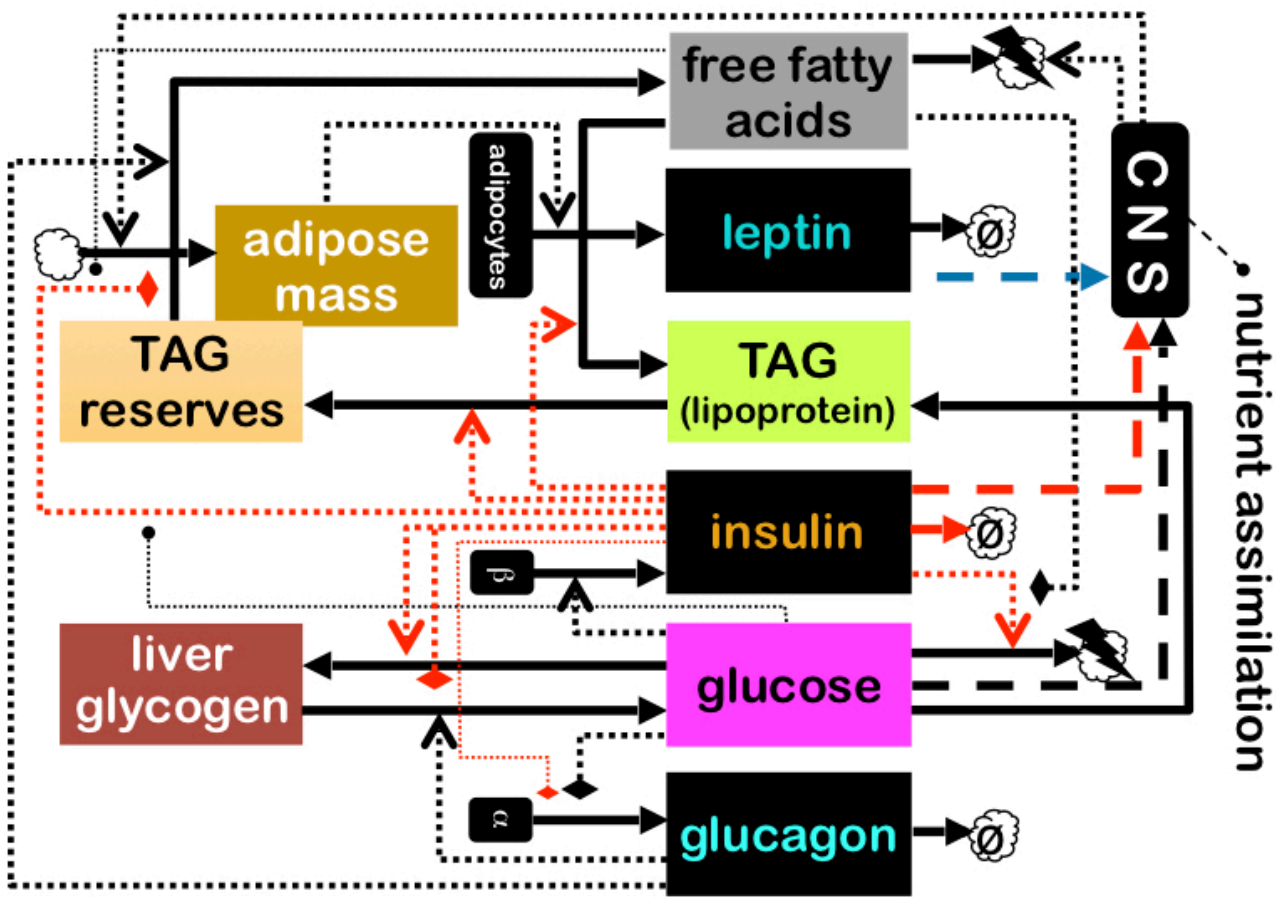


Figure 4

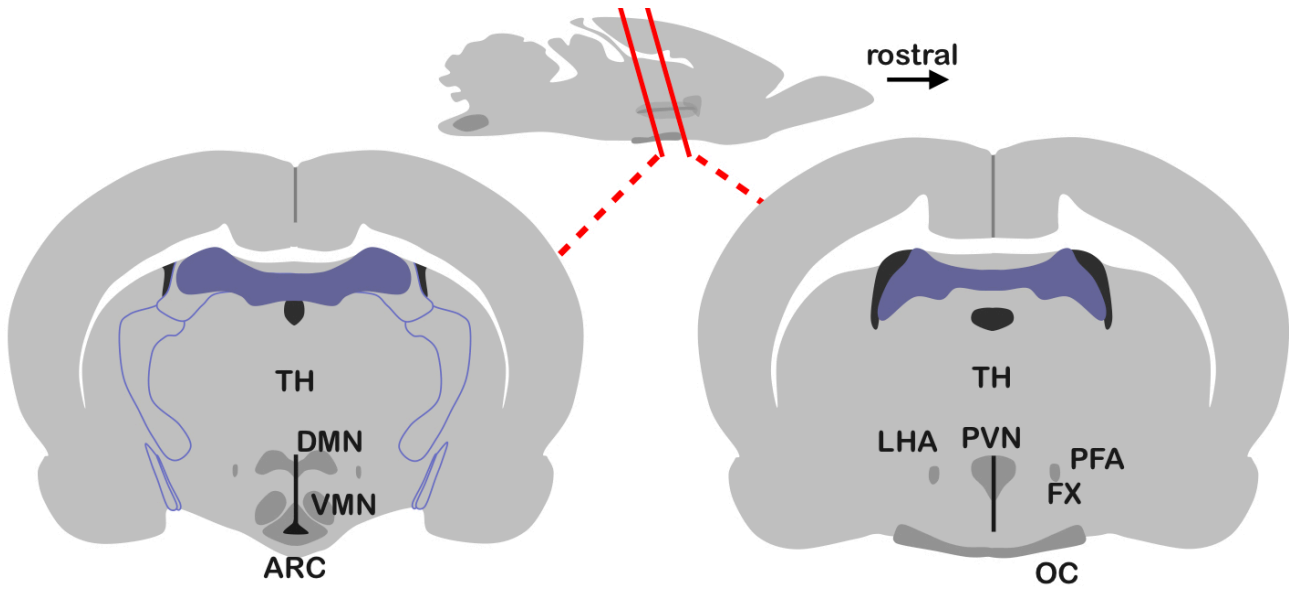


Figure 5

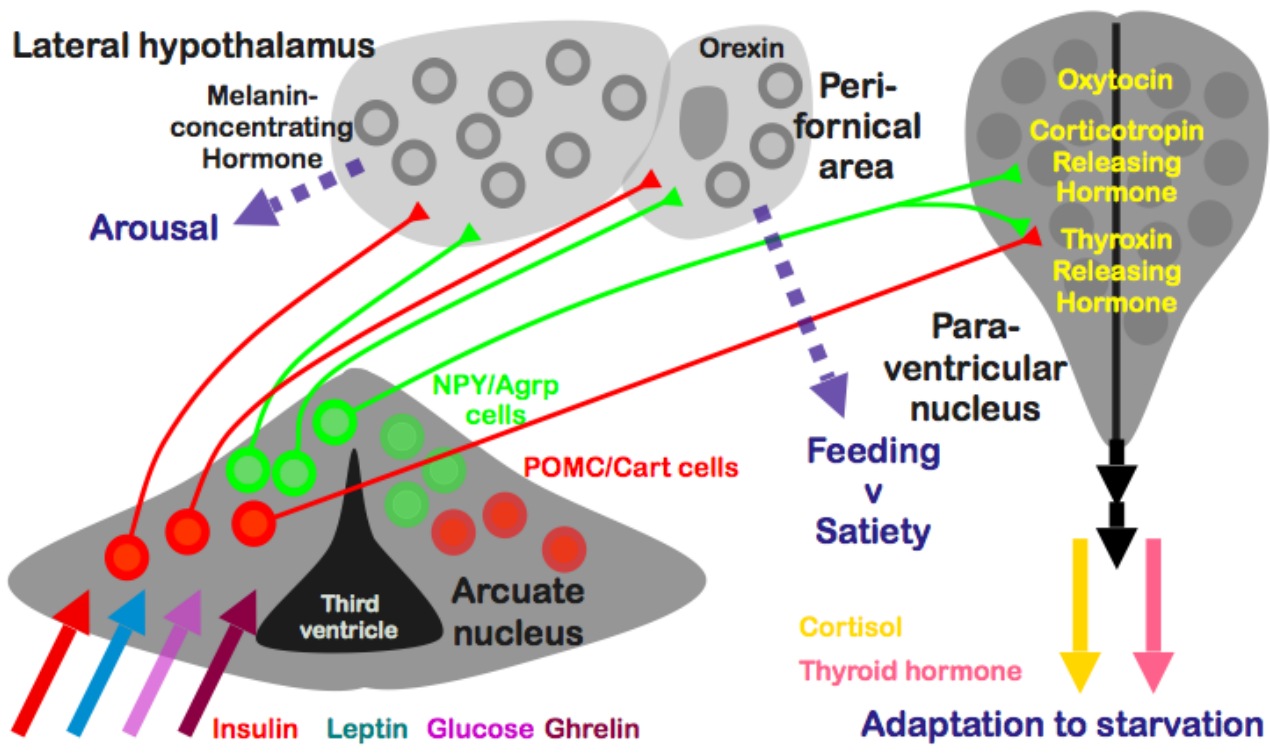


Figure 6

Figure 7

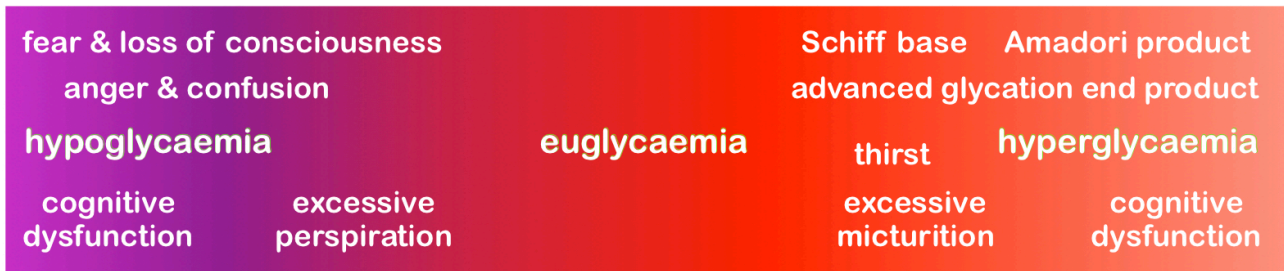


Figure 8

