Systems Medicine BE333 Lecture notes

Uri Alon (Fall 2021)

Lecture 3

The stress-hormone axis as a two-gland oscillator

Introduction

So far, we discussed an endocrine gland controlling its own size, with beta-cells as our main example. We now explore what happens when *two* glands control each other's sizes. We will see that this creates a feedback loop with a timescale of months, which can produce noise-driven oscillations. We will explore such feedback in the HPA axis, a stress-response pathway whose output is the hormone **cortisol**. Cortisol is a hormone that prepares our body for physical and psychological stress.

The HPA pathway is clinically important: it is involved in problems of chronic stress like high blood pressure, and in mood disorders such as major depression and bipolar disorder. Cortisol analogs are important drugs, used to suppress inflammation and prevent tissue swelling. These drugs are glucocorticoid steroids, often simply called **steroids**.

We will see how the two-gland feedback loop

- Creates seasonal hormone oscillations that provide a hormonal set-point for each season.
- Explains clinical phenomena with month's timescale, such as the side-effects of steroid withdrawal.
- Offers a possible origin for the timescale of mood fluctuations that last months in bipolar disorder.

The HPA axis controls responses to physical and psychological stress

When we wake up, we get a surge of energy (for some of us, including the author, this takes a bit of time), thanks in part to a morning surge in the blood concentration of the hormone **cortisol**. We get a similar surge of cortisol if we see a threat, such as a truck hurtling toward us. The surge focuses the mind and gets us ready to run or fight. We get a cortisol surge even if we *predict* that a threat is coming, such as a serious psychological stress. The hormone cortisol takes minutes to act, and a surge lasts for about 90 min. Stresses also activate much faster responses within seconds. These fast reactions act through the sympathetic nervous system and secretion of adrenalin that makes our hearts beat fast and the liver secrete glucose (which we won't

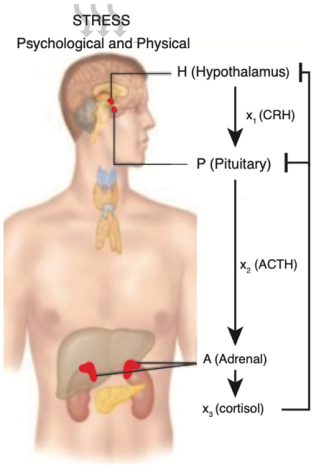


Figure 6.1

discuss in this lecture). Without cortisol and adrenaline, we wouldn't have the energy to stand up.

The hormone system that controls cortisol is composed of a series of three glands, which we will denote H, P and A. It is thus called **the HPA axis** (Fig 6.1). H stands for the hypothalamus, a brain region between our ears that receives neural inputs from many other brain regions and uses this information to regulate hormones that act on the rest of the body. H is activated by emergencies and stresses, including low glucose, low blood pressure, inflammation, and psychological stresses. It is activated by alcohol and other addictive substances, a fact which will play a role in an upcoming lecture on addiction. It is also regulated by the brain's circadian clock which keeps track of the time of day, to generate the morning surge of cortisol and low cortisol at night.

All these inputs are processed in H and combined[1] into the output of H: secretion from H of the hormone CRH, which we will call x_1. This hormone is secreted into a private blood vessel, called the portal system, which flows into the next gland P, a pea-sized organ at the bottom of the skull called the **pituitary**.

In P, the hormone x_1 stimulates cells to secrete the hormone x_2 (ACTH) into the circulation. Together with x_2, P also secretes beta-endorphin, which is a painkiller and causes euphoria (endorphine means 'endogenous morphine').

The messenger hormone x_2 flows with the blood and reaches the third gland of the axis, A, the adrenal cortex. The word 'cortex' refers to the outer layer of the adrenal gland[2]. There are two adrenal glands, pyramid-shaped tips on top of the two kidneys. The cells in A secrete the final hormone, cortisol, denoted x_3. Cortisol closes a negative feedback loop by inhibiting the rate of production and secretion of its two upstream hormones, x_1 and x_2 (Fig 6.1).

Cortisol is a small fat-like hormone (steroid hormone). Because it is fatty, it can penetrate the membranes of all the cells in the body. It is sensed by special receptors inside the cells. When these receptors bind cortisol, they go into the cell nucleus, bind the DNA at specific sites, and cause the expression of genes that respond to stress. Cortisol has many effects, including increasing the production of glucose, increasing blood pressure, and suppressing inflammation. It also has effects on attention and memory. Cortisol also diverts resources from projects like reproduction and growth towards immediate action. It gets the body ready to respond to the stress, and to prepare for more stress.

Sudden stresses cause an **acute** response, in which cortisol levels rise from their normal range of 100-300 nM to nearly 1000nM. These pulses last about 90 min, the half-life of cortisol. If the stress lasts for weeks or more, as in psychological worries, cortisol causes the symptoms of **chronic stress**: weight gain, high blood pressure, risk of heart problems and diabetes, bone loss and depression. Chronic high cortisol also causes cognitive changes, including increased sensitivity to negative stimuli and heightened anxiety, as well as decreased learning (technically, issues include decreased inhibition of the amygdala by the prefrontal cortex, and damage to the hippocampus).

---

[1] How the inputs are combined is an open question: is it a product or a sum of the inputs? It is likely to be a product.

[2] Incidentally, the middle of the adrenal gland secretes (and gives its name to) adrenaline. It is not part of the HPA axis.

That is one reason that we take nice deep sighs of relief in our class. Deep breaths relax the fast sympathetic nervous system for **fight & flight**, and activate its counter system called the parasympathetic nervous system for **rest & digest**. They cause a nice perception of the class as a pleasant place, which reduces stress. The HPA axis is in fact implicated in mood disorders, as we will discuss later in this chapter.

Cortisol for optimal functioning needs to be in a middle range. Insufficient levels result in low energy, whereas excessive levels over long times cause the symptoms of chronic stress (Fig. 6.2). The middle range of inputs to the HPA axis can be thought of as healthy stimulation, as in sports and challenging activities that interest us. The lower range is weakness and boredom and the higher range excessive stress.

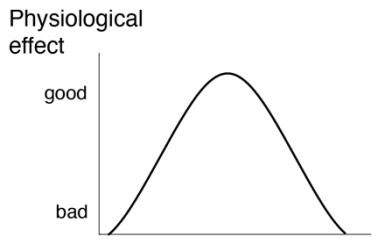


Figure 6.2

(*Song The stress pathway, HPA*/music psychokiller, Talking heads

I can't seem to face up to the facts,

tense and nervous, can't relax,

can't sleep, heads on fire

don't stop me I'm a real live wire.

The stress pathway, HPA, fafafafafafafafafa

Fight freeze or run run run away*)

The classic model for the HPA axis explains responses on the fast timescale of hours, but not on the slow timescale scale of months

The classic description of the HPA axis discussed so far is what you find in textbooks. It's the "HPA version" of the minimal model we saw for glucose and insulin. Let's write it down mathematically so we can explore its properties. Our main conclusion will be that the model shows the expected response to acute stress but cannot explain phenomena on the scale of weeks-months. We will later expand the model by adding control of gland sizes and open a world of phenomena on the month's timescale.

We lump together all the stress inputs into a single quantity u(t). This input u causes H to secrete the hormone x1, which is degraded with a timescale of minutes at rate α_1. This can be described as follows, using the same logic as we did for the glucose minimal model of lecture 3:

$$(1)\ \frac{dx_1}{dt} = q_1\ \frac{H\,u}{x_3} -\ \alpha_1\, x_1.$$

the inhibitory effect of x_3 on the secretion of x_1 is described[3] by the term $1/x_3$, so that the more cortisol x_3, the less production of x_1. The removal rate $\alpha_1$gives x_1a halflife of minutes.

The hormone x_1 acts on pituitary cells (called corticotrophs), whose total mass is P, to induce secretion of the second hormone x_2. Secretion is inhibited by x_3. The hormone x_2 has a lifetime of $1/\alpha_2 = 10$ minutes, and so

$$(2)\ \frac{dx_2}{dt} = q_2\ \frac{P\ x_1}{x_3} - \alpha_2\ x_2$$

The hormone x_2 in turn induces the cells of the adrenal cortex A to secrete x_3, whose lifetime is $1/\alpha_3$=90min:

$$(3)\ \frac{dx_3}{dt} = q_3\ A\ x_2 - \alpha_3 x_3$$

To solve for the steady-state hormone levels, we set temporal derivatives to zero, d/dt=0, as in the previous chapters, and do the algebra. The cortisol steady state depends on all of the model parameters, and rises with the input u and the sizes of the two glands A and P

$$(4) x_3^{st} = \left(\frac{q_1 q_2 q_3}{a1a2a3}\ u\ HPA\right)^{\frac{1}{3}} \sim u^{\frac{1}{3}}\ P^{\frac{1}{3}}\ A^{\frac{1}{3}}$$

Similarly, the steady states of the other two hormones also depend on all parameters. x_2 rises with the size of P but drops with the size of A, because A inhibits its upstream hormone x_1

$$(5)\quad x_2^{st} \sim u^{\frac{1}{3}} A^{-\frac{2}{3}}\ P^{\frac{1}{3}}$$

$$(6)\quad x_1^{st} \sim u^{\frac{2}{3}}\ P^{-\frac{1}{3}}\ A^{-\frac{1}{3}}$$

These equations correctly show that the hormone levels rise when stress input u occurs, with a timescale of minutes-hours.

These equations, however, do not explain phenomena on the scale of weeks-months:

(i) Effect of taking steroids to suppress inflammation – if taken for weeks or more, these drugs shouldn't be stopped at once. If they are, the body is unable to produce enough cortisol. Steroids need to be gradually tapered down, to avoid risk of dangerously low levels of cortisol.

(ii) The month-timescales of depression and bipolar disorder.

Moreover, the glands A and P are made of cells that divide and are removed (the H gland is in the brain and, like most brain neuronal tissues, it does not show sizable turnover of cells. Most neurons are not renewed, with some exceptions such as certain neurons in the hippocampus). To

[3] Binding of cortisol affects production as $1/(\mathrm{K}+\mathrm{x3})$ where K is the binding affinity to the receptor- called the Michalis-Menten equation explained in Appendix A. The $1/\mathrm{x3}$ term is a good approximation when x3 concentration exceeds K, as it does for cortisol whose blood levels are on the order of 100nM, and K is on the order of 10nM for the low-affinity mineralocorticoid receptor. Another cortisol receptor with a high K, the glucocorticoid receptor, comes into play at very high cortisol levels, and we will deal with it when we discuss depression.

maintain proper **organ size**, the A and P glands must have feedback control as discussed in lecture 4 on the BIG model. Since the glands act in a series of hormone secretions, it turns out that they control each other's growth. This situation is described next.

The HPA axis shows a feedback loop in which two glands control each other's size

The cells of the A and P glands proliferate and are removed with a typical turnover time of about 1-2 months. They have feedback loops that balance cell proliferation and removal. The inputs to these loops are the HPA hormones themselves. The hormone x_1 acts not only to induce secretion of x_2 by P cells; it also increases the P cell proliferation rate. Similarly, x_2 acts to induce x_3 secretion and to induce adrenal cortex A cell proliferation (Fig 6.3).

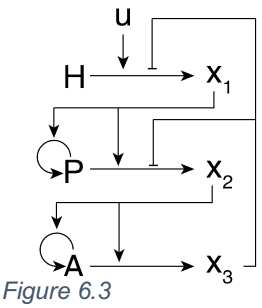


Figure 6.3

Thus, **the HPA axis shows two occurrences of the BIG circuit motif we saw for the insulin-glucose system**, in which a signal controls both secretion and proliferation. In the HPA axis, the two motifs are stacked on top of each other.

To reflect for a moment, in this course we learned quite a lot: a few weeks ago, it would have been difficult for you to make sense of such a complex circuit; but we will now see that in about ten minutes you will understand much of its behavior.

The equations for P and A total cell mass are the difference between proliferation and removal

$$(7)\ \frac{dP}{dt} = P(b_P x_1 - a_P)$$

$$(8)\ \frac{dA}{dt} = A(b_A x_2 - a_A)$$

Here a_P, a_A are the cell removal rates ($1/a_A$,$1/a_P \sim 1-2$ months), and b_P, b_A are the hormone-dependent proliferation rates.

These two equations have important consequences for the system on the scale of months. Here we again use the concept of **separation of timescales**. There are two different timescales in the system, a fast timescale of hours for the hormone production and removal, and a much slower timescale of months for changes in gland masses. Therefore, the steady states from Eq (4-6) correctly describe what happens if the stress input u is present for a few hours. But if u is present for weeks, the masses of the glands A and P increase gradually, and the system finds its full steady state with an enlarged adrenal cortex (Fig 6.4). Thus, Eq 4-6 are more accurately described as '**quasi-steady-states**'.

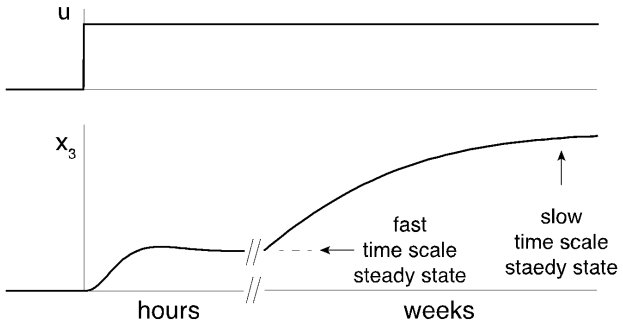


Figure 6.4

When we solve the steady-states of the 'slow equations' Eq 7-8, we find that the steady-state hormone levels become much less sensitive to the physiological parameters of the HPA axis. In particular, the only way to get to a steady-state (with A,P>0) is $x_{1,st} = a_P/b_P$, $x_{2,st} = a_A/b_A$.

These steady-state hormone levels therefore do not depend at all on production parametersq_1, q_2, q_3, removal rates etc. Cortisol similarly turns out to have a simple steady-state level robust to almost all model parameters (by plugging in x_1^{st} to Eq 1):

$$(9) x_3^{st} = \frac{q_1 b_P}{\alpha_1 a_P} u$$

Nicely, cortisol rises linearly with the input u, as it should. It provides a proportional response to stress inputs. This is an example of **allostasis,** in which the setpoint of a circuit can change according to signals. Indeed, cortisol setpoint can be high in chronic stress or chronic illness due to high input u from the hypothalamus. Contrast this plastic set point with the strict homeostasis of glucose, which in health does not stray from its 5mM setpoint.

Cortisol steady-state concentration does not, however, depend on many of the model parameters. This can eb described therefore as "robust allostasis". For example, the setpoint does not depend on the production parameters of the pituitary and adrenal, q_2, q_3 which vary, for example, with blood volume and metabolism. It also does not depend on the hormone removal parameters α_2, α_3. As in the case of glucose, the circuit 'protects' cortisol baseline levels from variations in many physiological parameters, compensating these changes by means of slow growth and shrinkage of the glands.

The gland mass changes also protect the dynamics. Cortisol dynamics after an acute stress input are independent on q_2, q_3, an example of dynamical compensation that we saw in the previous chapter for glucose.

We can derive equations for the gland masses on the timescale of months by using the quasi-steady state approximation for the hormones. This means using the fact that on the timescale of months, hormones can be assumed to be at their fast-steady state levels, Eq 4-6. Plugging x_1 and x_2 into Eq 7-8, shows a **negative feedback loop between P and A**:

$$(10) \frac{dP}{dt} = P(c_P u^{\frac{2}{3}} P^{-\frac{1}{3}} A^{-\frac{1}{3}} - a_P)$$

$$(11) \frac{dA}{dt} = A(c_A u^{\frac{1}{3}} P^{\frac{1}{3}} A^{-\frac{2}{3}} - a_A)$$

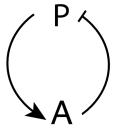


Figure 6.5

P acts to increase the size of A, by the action of x_2 (Fig 6.5). In contrast A acts to shrink P, because the hormone it secretes, cortisol, reduces x_1 which is the growth factor for P.

The P and A glands form a negative feedback loop that can oscillate with a seasonal timescale

For the remainder of this lecture, we focus on the timescale of months. This is the timescale of the negative feedback loop between the two glands. Like many negative feedback loops in

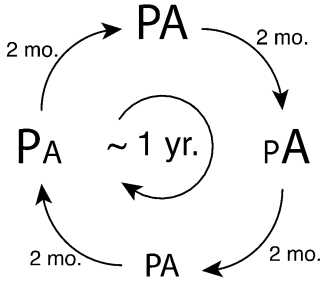


Figure 6.6

which both arms have similar timescales, the HPA circuit can show damped oscillations. An intuition for how delays in feedback can give oscillations can be found in old showers, like the shower I had growing up. The water would start cold, and I would turn it to hot, but due to the delay, I'd go too far, and the water would be scalding. So I'd turn it back to cold and then after a delay it would be freezing, and so on in a cycle of argh, ouch, argh..

To see how the oscillations go, follow Fig 6.6 from the top and go with the arrows, clockwise. When A is large it reduces P (first arrow). Smaller P reduces A. Smaller A makes P grow back. Large P makes A grow, closing the cycle (Fig 6.6). Each of these steps has a delay due to the tissue turnover time, about 2 months. The overall time for a full cycle is **on the order of a year** .

When Avichai Tendler in his PhD with me found this timescale of a year, we got goosebumps. A year is important for organisms because of the changes in seasons. It is useful to have **an internal clock that can keep track of seasons**, as we will discuss below.

To understand this clock, it helps me to make one more analogy of the A-P feedback loop to a *mass on a spring*. Suppose the mass is at rest at its equilibrium point (Fig 6.7). If you pull the mass down, the spring will stretch and cause a force pointing back to the equilibrium point. If you let go of the mass it will return to equilibrium but with high velocity v, so it will overshoot and compress the spring. The compressed spring pushes the mass back down, and so on, to obtain sustained oscillations (Fig 6.7).

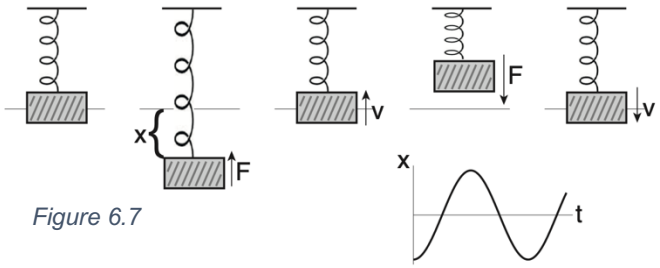


Figure 6.7

To describe this motion, we can use Hooke's law of the spring: the force grows with the extension of the spring, $F = -kx$. This force, according to Newton's law, F=ma, changes the acceleration *a,* which is the rate of change of velocity $a = dv/dt$. Thus, spring extension x 'inhibits' velocity $dv/dt = -\frac{k}{m}\, x$. Conversely, velocity 'activates' spring extension since by definition velocity is the derivative of position $dx/dt = v$. This is like a negative feedback loop in which v enhances x but x inhibits v (Fig 6.8). Thus, x and v are like A and P, and the negative feedback can act as an oscillator.

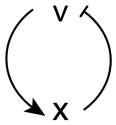


Figure 6.8

When you simulate the slow equations for A and P you don't get sustained oscillations; instead, you see damped oscillations that decay away. To get damped oscillations in the spring analogy, we can submerge the spring in a container with a viscous fluid like honey. The honey causes a **drag** force which is proportional to velocity: $dv/dt = -kx - b\, v$ (Fig 6.9). The spring does not oscillate forever but instead settles down to its equilibrium point.

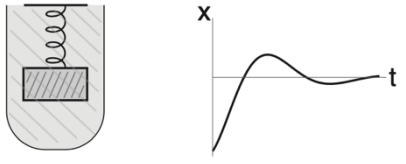


Figure 6.9

There is also a "drag"-like force in the equations for A and P. For example, large A makes more cortisol, which inhibits x2, the growth factor of A. Thus, large A acts to reduce its own growth,

similar to a large velocity v which causes more drag and slows itself down (thanks to Z. Tan in the 2020 systems medicine class for this comment).

Model for steroid withdrawal explains overshoot and recovery over months

When you pull the damped spring and let it go, it oscillates and relaxes to steady state. This provides a way to study the HPA oscillator, by seeing what happens when you pull it away from equilibrium and watch it recover. Let us therefore consider situations that perturb the HPA axis, reasoning that the two-gland feedback loop should show months-scale recovery dynamics from perturbations.

One common situation occurs when cortisol levels are high for weeks or more, and then are suddenly lowered. This occurs in the medical situation in which people take glucocorticoid steroids (long-lasting cortisol analogues such as dexamethasone or prednisone) for extended periods and then go off the drugs (Koch-Weser and Byyny, 1976; Dixon and Christy, 1980). Glucocorticoid steroids are given to reduce inflammation or suppress immune responses, as in asthma and autoimmune diseases, and after transplants.

Giving high doses of steroids for a few days usually causes no problems. But if they are given for two weeks or more, it is important not to stop steroid treatment all at once. If you do, the patient will show dangerously low cortisol. They can't make enough cortisol by themselves, causing serious symptoms: blood pressure drops to potentially fatal levels. This is called **steroid withdrawal**. Thus, one must gradually reduce doses over months- one protocol is to reduce the dose by 25% every two weeks.

Steroid treatment can be modelled by adding external (exogenous) cortisol D to the equations (by replacing x_3 by $x_3 + D$ in Eq 1 and 2). In this case, both pituitary and adrenal gland sizes shrink, due to inhibition of their growth factors x_1 and x_2 by D. It's as if the glands 'think' there is too much cortisol and shrink to return it to baseline. This causes the atrophied and involuted adrenal gland observed in extended glucocorticoid treatment (Nicolaides *et al.*, 2000).

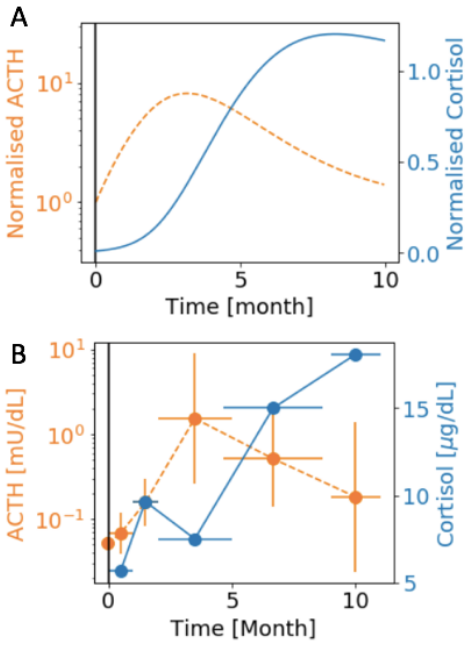


Figure 6.10

Thus, simulations of steroid withdrawal begin with small gland sizes A and P, after a long period of high exogenous cortisol D. They begin by setting D to zero. This is like pulling the spring hard, and then letting it go. You can see this in Fig 6.6, starting at the bottom (small A and P), and following the arrows. The dynamics show that P recovers first, and A follows it. As a result, x_2 overshoots after 3 months and goes back to normal after 9 months, whereas cortisol is abnormally low and recovers more slowly after 9 months (Fig 6.10A). The overshoot of x_2 is due to the release of x_1 inhibition, which causes P to grow. Only when P returns to normal size is x_2 sufficient to allow A to grow and recover.

Such overshoot dynamics were found by Graber et al, in an example of a "small-data" study which followed n=14 patients who went off prolonged steroid treatment or had a cortisol-secreting tumor in their adrenal removed (GRABER *et al.*, 1965). Patients showed a large (10-fold) overshoot in x_2 after a few months, and a slower recovery of cortisol over 10 months (Fig 6.10B), in agreement with the model.

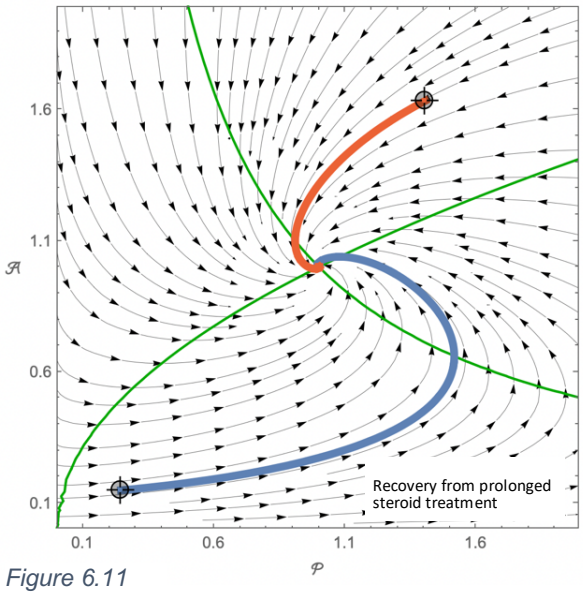


Figure 6.11

We can use the nullcline work we did in this course to get a nice phase portrait of the system on the scale of months (Fig. 6.11). The two nullclines (from Eq 7,8) are dA/dt=0, a rising line in which a given P enlarges A, and dP/dt=0, a dropping line in which a given A reduces P. We can plot little arrows in the phase portrait to show how the two glands 'flow' back to their fixed point at the intersection of the nullclines. This is the first phase portrait in our course where the two variables have similar timescales. The flow spirals into the fixed point.

Prolonged steroid treatment causes both glands to shrink. Recovery thus starts at the bottom left and follows the thick blue trajectory: first a growth of P, and only then a growth recovery of A with an overshoot. Incidentally, no overshoot is seen in the red trajectory when the two glands start larger than the fixed point, as can occur after chronic stress. The recovery there is nearly straight back to normal, but still takes months. It's as if the pathway is 'made' to recover from certain natural situations like prolonged stress, and not from artificial ones like steroid drugs. (Work of Moriya Raz and Avi Mayo).

As a theoretical point, to enhance our skill with nullclines, note how the arrows that cross each nullcline are always either horizontal or vertical. On the rising nullcline, A does not change by definition (dA/dt=0), only P changes, and the arrows are thus horizontal. On the declining nullcline, P does not change (dP/dt=0), and the arrows are therefore vertical. Just knowing this allows us to sketch the arrows and gives a sense of the spiral flow around the fixed point. Plotting phase portraits is easily done for any equations using online codes. I love the look of this phase portrait. See how steroid withdrawal is easy to read out (blue trajectory) by following the arrows: first P grows, then A, and then a spiral back to the fixed point.

Seasonality in hormones

With the mass-on-a spring-spring analogy we can also understand what happens to the HPA axis when it gets inputs that change over a year, like the length of the seasons **entrain** the oscillator to a period of one year. It's like putting the honey container on a platform that oscillates up and down: the spring picks up the platform's frequency (one year) and begins to oscillate with the same frequency. Even if the oscillator natural frequency (perhaps 10 months) is a bit different, it can still entrain effectively to a yearly input.

Due to this year-scale frequency, the two-gland feedback loop can be entrained (driven) effectively by signals that vary over the year. This includes daylight, day-length, temperature, and other environmental factors.

Indeed, animals have seasonal behaviors and physiological changes across the year. Birds and mammals migrate, change their feathers/fur, hibernate and mate in certain seasons. The signals for this seasonal cycle have to do with day length variation. day. Short days (short photoperiod) appear to be a kind of stress input. Days are shortest in December 21 and longest in June 21 (Fig 6.12). Day length affects a hormone secreted in the brain, called melatonin, which feeds into the P gland.

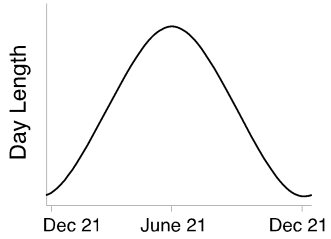


Figure 6.12

But animals do not rely only on external signals such as day-length to figure out which season it is. **Animals have an internal clock that keeps track of seasons**. Such an internal clock is crucial, for example, to avoid mistaking a series of cloudy days in summer as winter. Evidence for such a clock was found when various animals were kept in constant day-length and temperature conditions for years. They still showed their normal seasonal changes, although not with a 12-month period, but with a 10-month period (Fig 6.13). This inner clock is thus called a **circa-annual clock** (circa=about, annual=year). The mechanism for this seasonal clock is not well-characterized, but the pituitary is thought to be involved, due to its response to a day-length-dependent melatonin signal (Wood and Loudon, 2018).

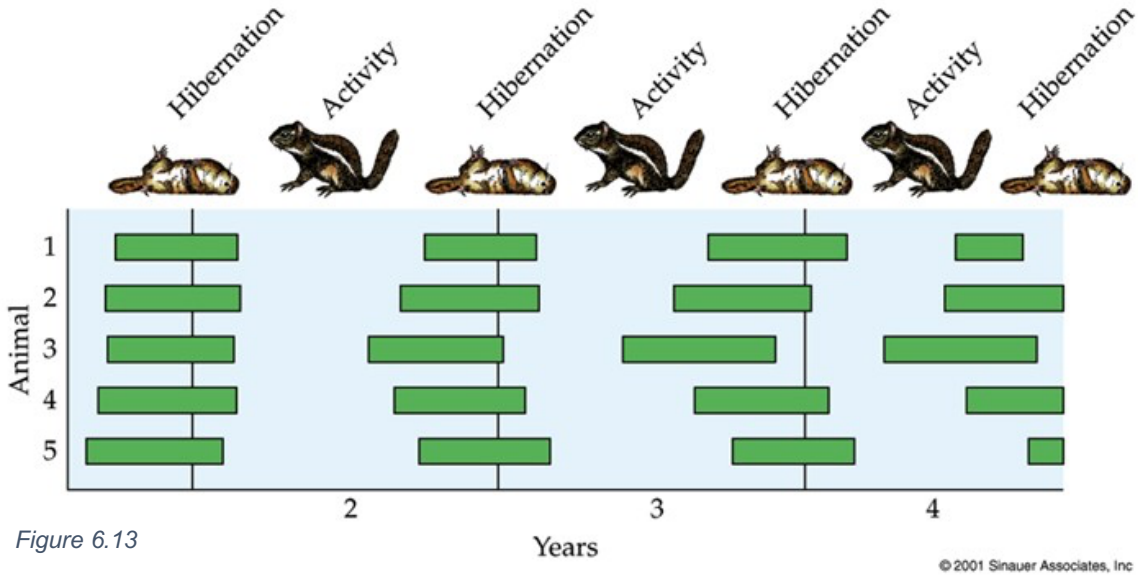


Figure 6.13

The long timescale of the $A - P$ oscillator may provide some of the inertia of this internal clock. A yearly oscillating day length input can drive the oscillator to show sizable seasonal oscillations. As shown towards the end of the lecture, noise can cause it to perform full, undamped oscillations with a natural frequency of about a year even without entrainment; this may explain the animal experiments kept in constant daylength and temperature conditions, and indicate that the cogs of the circannual clock are the P and A gland masses.

To study the effects of entrainment by yearly cycles of day-length, we can study the HPA axis with inputs that vary with a period of one year, $u(t) = 1 + u_0 cos\,(\omega t)$. The peak input occurs on the shortest day of the year, December 21. It is worth noting that the seasonal input depends on geography: The amplitude u_0 increases with latitude. It is about 2 times larger in London than in Tel Aviv (Teets, 2003).

A model with no gland-mass changes would predict that hormones simply follow the day length input. When input is highest (Dec 21 for cortisol), all hormones are highest. Adding gland mass changes, however, modifies this picture and provides specific predictions (Fig 6.14):

(i) Peak cortisol should not occur at the time of peak input (December 21) but rather be delayed by about two months. This delay is due to the delay in the feedback loop and is determined by only two parameters: The P and A tissue turnover times a_A and a_P. The model therefore predicts a cortisol peak at february-march, late winter/spring. The shift and amplitude can be derived analytically from linearized equations.

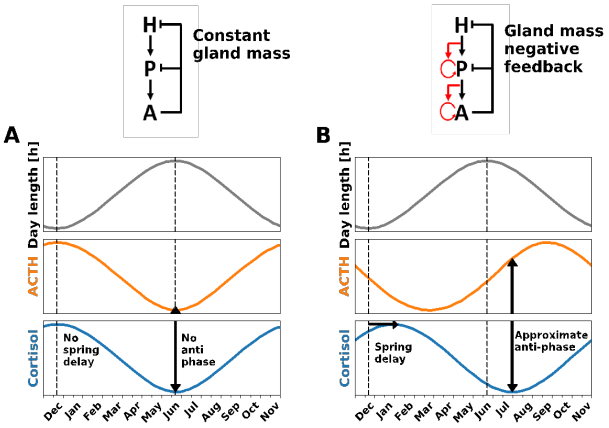


Figure 6.14

(ii) Cortisol and x_2 (ACTH) are predicted to be in approximate **anti-phase**, with x_2 peaking in late summer/fall. This antiphase cannot occur in a model without gland size changes, because in such models the two hormones vary together within minutes-hours of each other.

Large-scale medical records show the predicted oscillation phases

To test such predictions requires data on hormones collected in different months. Here we can see two types of data sources in systems medicine: hypothesis-driven 'small-data' and exploratory 'big-data'.

A small-data study is usually done with a focused research question and scope. An example is the UK Whitehall study on cortisol levels of thousands of British civil servants. One striking finding is that the lower the civil service rank, the higher the cortisol level. Indeed, socioeconomic status is a strong predictor of health and longevity across cultures: the lowest decile has more chronic stress, and has a mean lifespan shorter by a decade than the top decile.

When tested for seasonality, saliva cortisol levels (meta-analysis with n=18,698 subjects) was maximal in winter-spring and minimal in fall (Miller *et al.*, 2016). There was a 17% peak-to-trough amplitude between spring and fall saliva cortisol. A Swedish study of a dozen people found a saliva cortisol maximum in spring with a larger amplitude of 30% (Persson *et al.*, 2008), as predicted by the northern latitude of Sweden.

In addition to these specific studies on cortisol, one can utilize 'big-data'. An exciting advance in systems medicine is the increased availability of large, searchable medical datasets. These are

called **electronic medical records (EMR)**. For example, Israel's largest health insurer (Clalit) has agreed to allow researchers to study anonymized data on about half of the Israeli population over 15 years (2002-2017) totaling about 50 million life-years, with broad socioeconomic and ethnic representation. The data includes disease codes, drugs purchased, and blood tests. Thus, one can explore hypotheses or look for patterns. For seasonality, you can compare hormone tests according to the month of the year they were taken. This is the work of Alon Bar, a TA in our course, with Avichai Tendler and Avi Mayo, also a TA, in collaboration with Amos Tanay's lab.

A major issue with medical record datasets is **ascertainment bias**. The medical tests are done for a clinical reason, as opposed to a uniform sample of the population as attempted in many 'small-data' experiments. To address this bias requires ways to filter out data from people with medical conditions that can confound the results. For the purpose of detecting seasonality, for example, one can remove all people with chronic disease according to international disease codes. One can also filter out, for each blood test, data from individuals that took a drug that affects that specific test. Finding such drugs can be done from the medical record data itself, by finding drugs that significantly affect test results on average. Finally, one can remove tests which include indications for infection to avoid effects of yearly infection cycles.

A second major issue for hormone tests is **circadian rhythm.** Cortisol blood tests done at different times of day give different answers, with highest cortisol in the morning. Thus, it is important to consider only tests done at a certain time of day to avoid the effects of circadian oscillations in the hormones.

After controlling for these confounders, we can now plot hormone levels as a function of the month of the year. A picture emerges (Fig 6.15). In both males and females, blood cortisol (n=104,255 blood tests) shows a peak in spring and a minimum in fall, as predicted. The timing of the peak and trough correspond to tissue turnover parameters of about two months. There are fewer x_2 (ACTH) tests in the dataset (a few thousand after filtering, mostly in females), and hence more noise. Yet it appears that x_2 peaks in summer in females; males have too few tests to make meaningful conclusions (Fig 6.15, orange line).

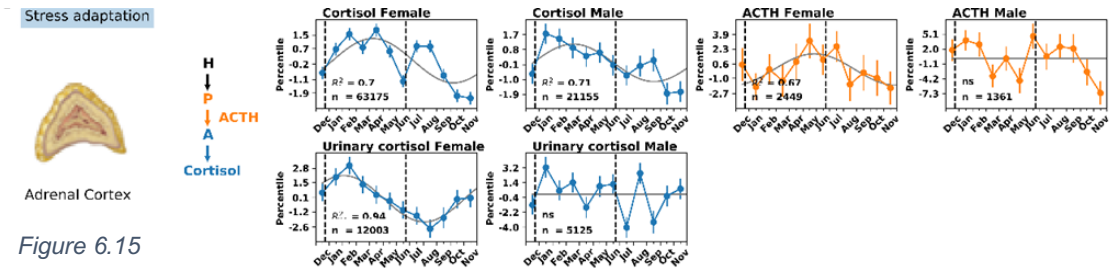


Figure 6.15

The oscillation amplitude is only a few percent. This matches the predicted amplitude at the $31°$ latitude of Israel. The model predicts larger amplitudes at higher latitudes (Fig 6.16), where daylength changes more strongly with seasons, reaching about 20% and 30% peak-to-trough in London and Stockholm respectively. Tests of

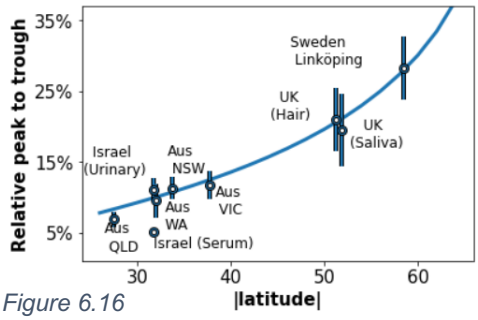


Figure 6.16

saliva cortisol in other countries show the expected rise with latitude (Fig 6.16). Tests in Australia show the right seasonal dependence but shifted by 6 months (Australian winter is in May-August).

Since the HPA axis is implicated in mood disorders (Watson and Mackin, 2006; Herbert, 2013), the seasonal variation in the HPA axis may contribute to **seasonal component of affective disorder (SAD)** (Rosen *et al.*, 1990; Mersch *et al.*, 1999; Avery *et al.*, 2001), also known as 'winter blues'. This is a relatively common disorder at high latitudes (10% prevalence in Alaska versus 1% prevalence in Florida). The spring cortisol peak might explain why the SAD symptoms peak in spring. For example, suicide rates peak in April (Bridges, Yip and Yang, 2005). Likewise, seasonal HPA oscillations can contribute to the seasonal component of major depression and other mood disorders.

Other HP axes control reproduction, growth, and metabolism, and show similar seasonal oscillations

The HPA axis is one of four axes that together control major fitness components of living organisms: reproduction, metabolism, growth, and stress. The hypothalamus H and pituitary P participate in all four. That is why P is sometimes called "the master gland". Each axis has its own x_1, x_2 and x_3 hormones (Fig 6.17). Just as in the HPA axis, H secretes x_1 hormones to stimulate pituitary cells to secrete x_2 hormones that stimulate an effector organ for each axis to secrete x_3.

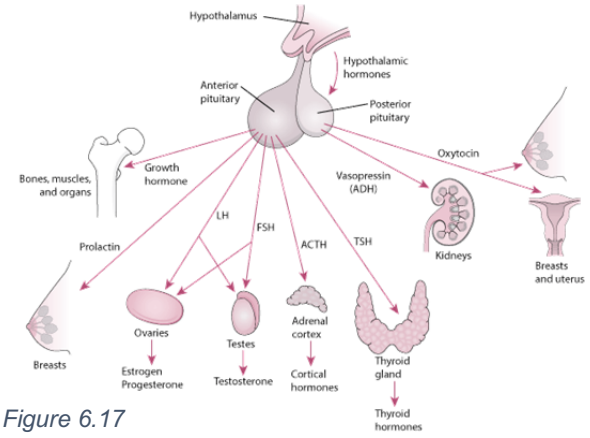


Figure 6.17

In the reproduction axis, for example, the x_2-hormones are FSH and LH, and the x_3 hormones are estrogen (from the ovaries in females) or testosterone (from the testes in males). For metabolism, the x_2 hormone is TSH, and x_3 is thyroid hormones secreted from the thyroid gland. For growth, x_2 is growth hormone, given as therapy for short stature as in the case of the child who became the football player Messi, and x_3 is IGF1. (Fig 6.18). These axes will be the stars of upcoming lectures.

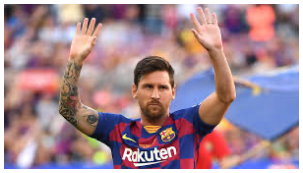


Figure 6.18

As in the HPA axis, these axes show a combination of two mass-control motifs stacked on top of each other (Fig 6.19). In all of them, there is a fast feedback loop in which x_3 inhibits secretion of x_1 and x_2. There is also a slow organ-size control feedback in which growth control feedback is at play. The hormones act as

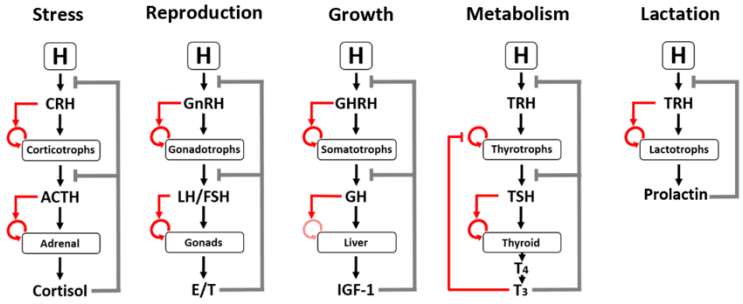


Figure 6.19

growth factors for the glands, with slight variations between the circuits in the configuration of the arrows (Fig 6.19).

This leads to the prediction that the x_2-hormones should peak in summer/fall and the x_3 hormones in winter-spring in all of these axes. Indeed, Clalit data shows that in both the thyroid axis and the reproduction axis, x_2-hormone blood test levels (TSH, LH, FSH) show a peak in late summer, about 1-2 months after June 21 (Fig 6.20/6.21). The x3 hormones (free thyroid hormone T3, estradiol, testosterone) show a late-winter/spring peak (Fig 6.20/6.21), as predicted. Oscillation amplitudes for most hormones are in the range of 4-10%. Some of these tests exceed 6 million data points(!), such as TSH tests, with error bars smaller than the dots in Fig 6.21. For all hormones see Fig 6.22.

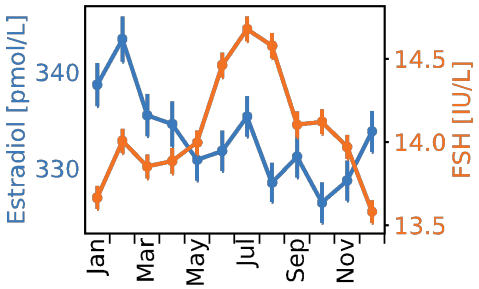


Figure 6.20

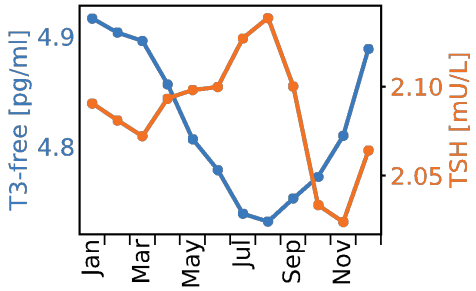
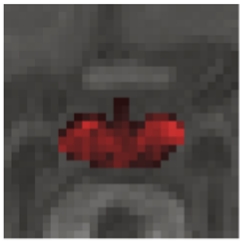


Figure 6.21

In all the axes, pituitary cells are predicted to be at maximal total mass in winter. The pituitary volume seems to change seasonally as predicted, based on MRI images (Fig 6.23)



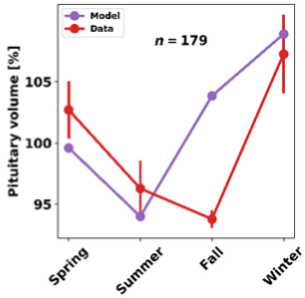


Figure 6.23

In contrast to the spring delay of the hormone oscillations, blood tests for key metabolites such as glucose, calcium and other important compounds and ions, show seasonal oscillations which peak around the shortest or longest days of the year with no delays. It seems that these metabolites lack a two-gland oscillator in their control systems.

Intriguingly, the late-winter peak of the different axes suggests that humans might have a peak season for reproduction, growth and HPA response. This is supported by data collected in studies that shows peak child growth, peak cognitive ability, and peak sperm quality in winter. Similarly, conceptions peak in winter in countries with latitude like Israel; peak conception shifts to summer in northern clines. Cultural and regional effects complicate the interpretation of conception seasonality. But overall, there is a case for a biological basis of human seasonality. Although

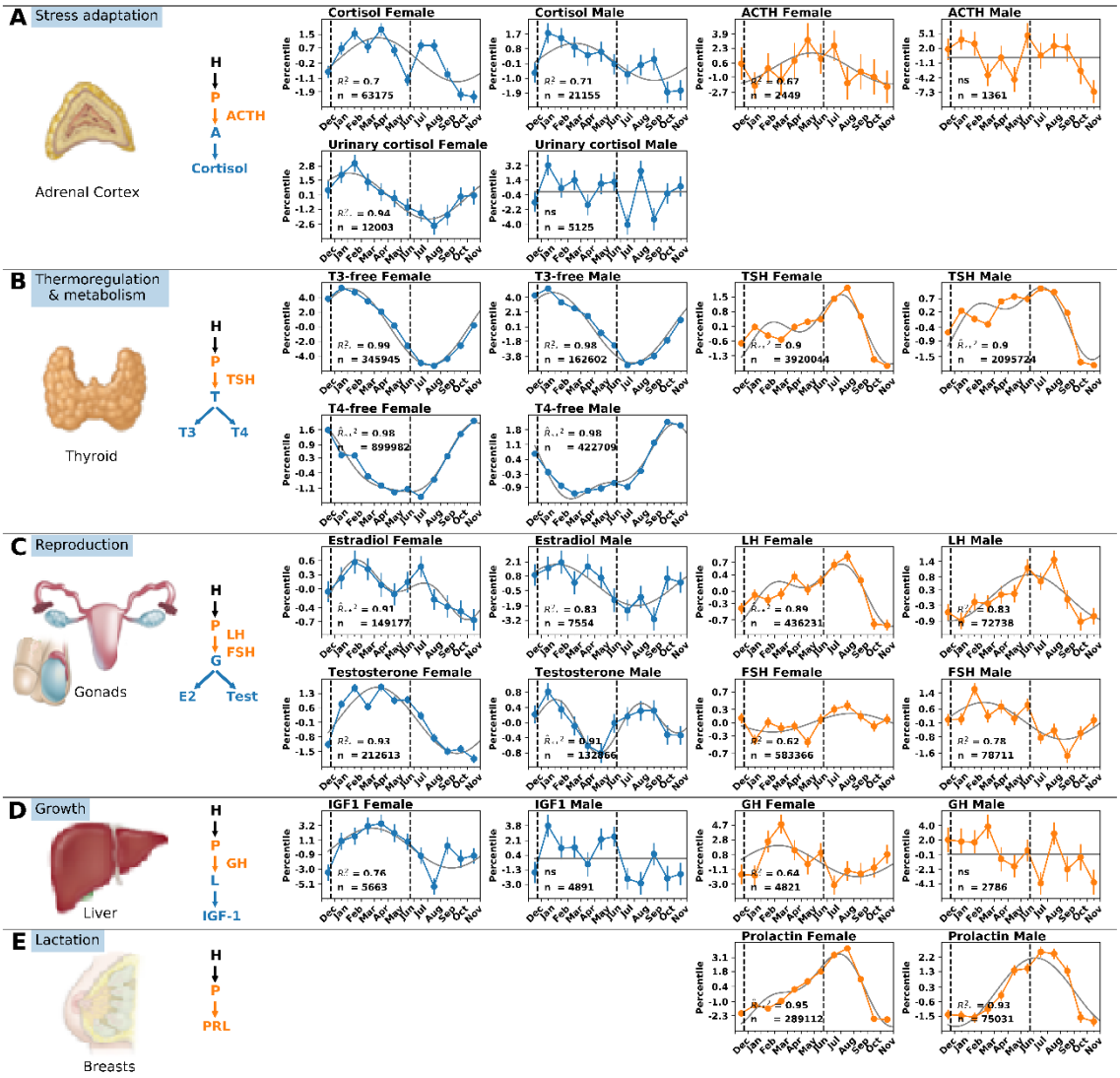


Figure 6.22

modernity makes seasons less impactful, in previous centuries and even today there is a season for every purpose.

Bipolar disorder has a timescale of weeks-months, and this timescale can be generated by tissue size fluctuations

Let's use these concepts to explore a mystery in psychiatry. We consider a major mood disorder called **bipolar disorder (BD)**. BD was formerly called manic depression. It is characterized by periods of months of depression with negative thoughts and low energy, and periods of mania for days to weeks with excessive high energy, irritability, and poorly-thought-out decisions. This is unlike familiar mood swings. Depression seriously interferes with the ability to feel joy, to function, eat and get out of bed; mania has delusions that can seriously damage relationships and work. BD is a leading cause of suicide and self-damage and causes suffering to about 1% of the worlds' population.

As in most psychiatric conditions, the biological understanding of BP is still lacking. It has a genetic component, but no genes of large effect have been identified. Medication such as lithium can stabilize moods for some people, but the mechanism is not clear. There is no blood test for depression or bipolar (or any other psychiatric condition so far), unlike diabetes where a glucose test is the basis for diagnosis. Diagnosis is done by interviews with psychiatrists. Since the mystery is so great, we can afford to speculate using the concepts of this lecture.

We focus on the timescale of BD: what sets a weeks-months timescale for mood swings? Neurons work in seconds, hormones in hours, and gene expression changes take a day or less. Few processes can supply such a month's timescale.

We can explore the hypothesis that the month's timescale is due to gland size variations. The glands of the HPA axis are natural candidates. The HPA axis is dysregulated in major depression (MD) and BD: about 50% of people with major depression have high cortisol and enlarged adrenal cortex. Conversely, high cortisol causes depression in a large fraction of individuals (as in Cushing's syndrome mentioned above). High doses of glucocorticoid drugs cause mania-like symptoms in some patients, and/or depression in others. Historically, the one of the first clues for the existence of a stress pathway came from work on suicide victims and stressed animals which showed enlarged adrenals. This was before anyone knew of hormones.

Often, BP episodes are preceded by stressful or joyous life events. To make a conceptual model for BP, we therefore need to schematically conceptualize psychological stresses. To do so, we simply note that life events lead to perceptions in the brain (Fig 6.24). Perceptions are person-specific and depend on experience and narratives. For example, a life event such as the victory of a political candidate in an election causes different perceptions in different people: some grieve, others rejoice, and others don't care. Perceptions affect our behavior which in turn affects life events. The HPA axis output, cortisol, also affects perceptions as discussed above.

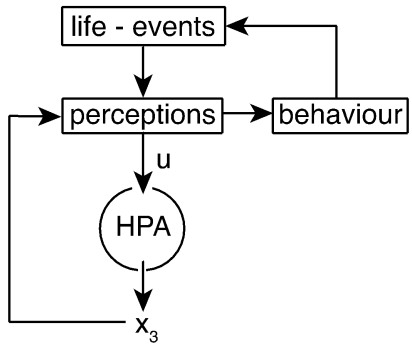


Figure 6.24

Let's suppose that people with BD tend to perceive life events in ways that create larger input signals to H, a kind of psychological excitability. Such excitability is often reported for people with BD. We can model these inputs, based on fluctuating perceptions, as a noisy input signal u(t) with a standard deviation sigma. In people with BD, let's assume that sigma is larger than in people without BD.

Such a noisy input to the HPA two-gland feedback loop causes a fascinating phenomenon: the loop shows noisy oscillations of gland sizes and in hormones, with a typical timescale of months. The larger the noise, the larger the amplitude of these oscillations. To see this, we can use the spring analogy: put the spring in its honey container on the back of a flatbed truck driving on a rough gravel road (Fig 6.25). The spring picks up vibrational frequencies close to its natural

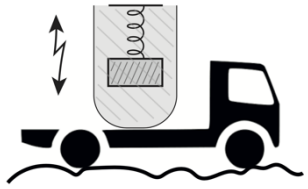


Figure 6.25

(resonance) frequency and starts making noisy, erratic oscillations (Fig 6.26). The rougher the road (more noise) the larger the oscillations.

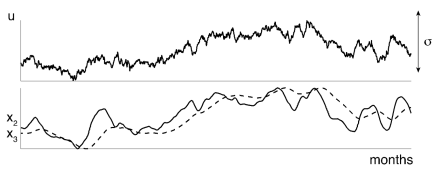


Figure 6.26

These noisy oscillations look like a messy ring around the fixed point in the phase portrait (Fig. 6.27, black trajectory). In fact, any system with a spiral fixed point - arrows that spiral into the fixed point instead of going directly into it - will have such noise-driven oscillations[4].

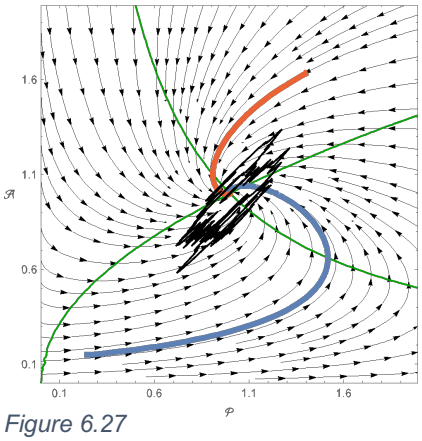


Figure 6.27

Note that these oscillations are due to stress inputs and are different from the seasonal oscillations discussed above. Seasonal oscillations are small. They ride 'on top' of the day-day stress responses. To see seasonal oscillations, you need to average over many people to average out their individual stress-responses. The noisy oscillations we are talking about now are much larger, and are not synchronized with the seasons; instead, they have a different phase in each person.

To test this, we did an experiment that used hair to measure cortisol over a year. Cortisol diffuses from the circulation into the hair follicle and sits passively in the hair (Fig 6.28), a phenomenon with no physiological consequence but a cool way to measure the hormone over time. Since each centimeter of hair is approximately a month of growth, a 12cm hair sample contains a record of the cortisol history over a year. Healthy participants showed fluctuations with periods of months to a year year, as predicted by the model (Maimon et al 2020). These fluctuations were not seasonal oscillations (peak was not in winter for all participants) and were much larger- amplitude of about 20% instead of the seasonal few percent. The experiment was small (55 participants), did not test the correlation between moods and cortisol, and needs to be repeated.

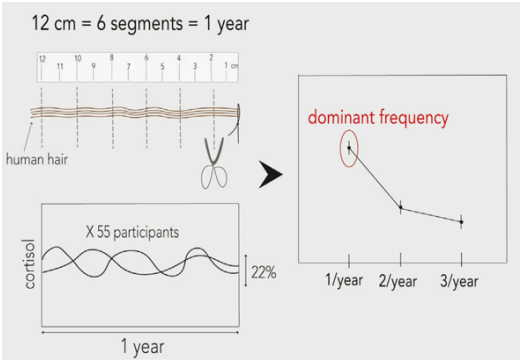


Figure 6.28

Thus, it is possible to hypothesize that the noisy oscillations of the HPA axis might account for some of the inertia and timescale of mood swings in BD (Fig 6.29). Similar oscillations in the HP-thyroid axis may cause energy ups and downs typical of BD. A question for current research is how (and whether) the hormones and gland sizes map to mania and depression. We are now repeating the hair experiments with participants with BD. Preliminary results from n=24 participants

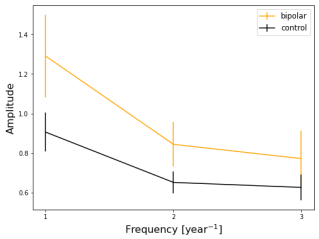


Figure 6.29

[4] For experts in linear analysis, the frequency of the noise-driven oscillations equals the frequency of the spiral, determined by the imaginary part of the eigenvalues at the fixed point.

with BD suggests larger relative amplitude of year-scale fluctuations than control participants, as predicted.

If such HPA oscillations turn out to be important for BD, to cause some aspects of BD and not merely be a readout, one might consider them as a pathway for treatment. One can build a controller- an algorithm that administers cortisol at the right phase of the oscillations, in a way that tends to dampen the oscillations out (Fig 6.30). This is like an effect known to anyone who ever rode on a swing: if you kick at the right time, you can gradually stop a swing from swinging. Indeed, simulations suggest that such a controller can reduce oscillation amplitudes by a large factor (Fig 6.30). These directions are a current research project in our group.

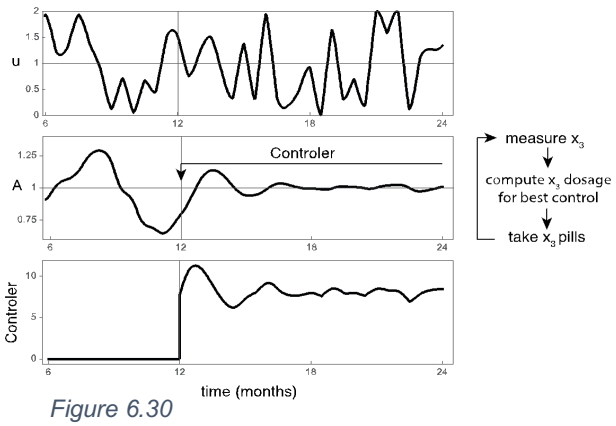


Figure 6.30

We might ask what purpose do moods serve in healthy physiology? A leading hypothesis is that moods evolved as a mechanism to allocate effort in proportion to reward (Medzhitov and Stearns, 2016, quoting Nesse,2015). “When payoffs are high, a positive mood increases initiative and risk taking. When risks are substantial or effort is likely to be wasted, low mood blocks investment.” There is thus evolutionary benefit to a system that can slowly change moods according to integrated information on risks and rewards. The system learns about risks and rewards over the timeframe of weeks and is not fooled by an outlier ‘positive’ stimulus into thinking that things are good. The inertia of HPA gland sizes can help provide such a swinging pendulum for moods. This creates a tradeoff, however, a fragility to mood disorders including depression and bipolar disorder. Since genes affect many aspects of behavior, and gene effects add up, there will be a distribution of set points between people: some people are more excitable, others more even keeled. Some people draw a combination of genes that lies at the extremes of the distribution, with increased fragility to disorders. The fragility can be exposed by stressful life events.

To sum up, when two organs control each other’s sizes, an oscillatory phenomenon on the timescale of months can occur. The two-gland oscillator allows the HPA stress pathway to synchronize with the seasons, with different hormonal set-points for different times of the year. It explains the need for prolonged tapering of steroid drugs after a long-term treatment. A cost of this oscillator may be noisy mood swings in a small percentage of people prone to BP. More generally, the body can be considered as an ensemble of interacting organs that constantly adjust their size and activity to adapt to changing conditions and to the states of the other organs.

(* Song/YMCA

It's fun to learn about the H, HPA

It's fun to learn about the H, HPA

Its got hormone swings for your every need

It gives you courage and it gives you speed,

It's fun to learn about the H, HPA

It's fun to learn about the H, HPA… *)

References:

Tendler A., Bar A., Mendelsohn-cohen N., Karin O., Korem Kohanim Y., Maimon L., Milo T., Raz M., Mayo A., Tanay A. & Alon U. "Hormone Seasonality in Medical Records Suggests Circannual Endocrine Circuits" Proceedings of the National Academy of Sciences, 118:2003926118 (2021)

Karin O., Raz M., Tendler A., Bar A., Kohanim Y. K., Milo T. & Alon U. "A New Model for the HPA Axis Explains Dysregulation of Stress Hormones on the Timescale of Weeks" Molecular Systems Biology, 16:209510 (2020)

Maimon L., Milo T., Moyal R. S., Mayo A., Danon T., Bren A. & Alon U. "Timescales of Human Hair Cortisol Dynamics" Iscience, 23:101501 (2020)

ROBERT M. SAPOLSKY **-** Behave **-** THE BIOLOGY OF HUMANS AT OUR BEST AND WORST

"Why zebras don't get ulcers" Robert M RM Sapolsky - Sapolsky (3rd ed.). New York, 2004

Smith SM, Vale WW. "The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress". Dialogues Clin Neurosci. 2006;8(4):383-395. doi:10.31887/DCNS.2006.8.4/ssmith

Belvederi Murri M, Prestia D, Mondelli V, Pariante C, Patti S, Olivieri B, Arzani C, Masotti M, Respino M, Antonioli M, Vassallo L, Serafini G, Perna G, Pompili M, Amore M. "The HPA axis in bipolar disorder: Systematic review and meta-analysis." Psychoneuroendocrinology. 2016 Jan; 63:327-42. doi: 10.1016/j.psyneuen.2015.10.014. Epub 2015 Oct 21. PMID: 26547798.