Notes of all papers read so far

Thejasvi Beleyur

Last Updated : 2020-09-01

Contents

What this book is about

This book is a compilation of all the notes I will be making for the papers that I read from now. I've been realising my paper notes are scattered everywhere, across multiple folders and multiple computers, this is my attempt at trying to unify everything into one place.

Chapter 1

Toledo et al. 2020, Science

Cognitive map–based navigation in wild bats revealed by a new high-throughput tracking system. (Toledo et al., 2020)

- *notes taken on 2020-07-14*

1.1 Introduction

- map based navigation goes beyond just simple modes eg. beacon following or landmark based navigation.
- bats are known to return to their normal sites even after displacement, which suggests 'map-and-compass' navigation style
- Authors' previous results showed that fruit bats flew straight paths, but this was limited to a few nights of data.
- In this study, authors managed to study 172 bats over a cumulative of 3449 nights.

1.2 Methods

- *ATLAS* - a reverse GPS system, where the animal wears a tag that emits a signal which is received by multiple ground stations - and thus using TOADs, the animal can be detected.
- ATLAS coverage region is ~88,200 hectares (or an area that's 29X29km big!!)
- bats tagged, and all fruit trees within a given region recorded.
- Also performed translocation experiments. Each bat was translocated to the periphery of its normal foraging area, but within detection range of their foraging area

- Also performed time-lag embedding to understand how complex the navigational mechanism is

1.3 Results

- Bats exhibited straight tracks, which is indicative of goal-directed behabiou

- Each bat had its favourite tree, and visited it every night, and even visited it from multiple directions of arrival

- Solid evidence for a cognitive map is when an animal moves between two points that can't be detected/seen/observed from each other (ie. it requires a kind of 'rigorous' mapping)

- 4.3% of all tracks,and 70/172 bats actually showed such shortcuts

- There was no difference in the rate at which shortcuts happened between the age groups of bats tagged

- *following a conspecific* – they talk about it by saying that in their dataset, they didn't see individuals flying close together - but they only tagged 172 bats of ??? thousand in the whole population

- Translocated bats were able to return to their normal foraging area

- Time-lag embedding showed a high-dimensional correlation (?) indicating there must be many difference navigational factors

  - If bats were following a simple navigational route, they might always arrive and depart from the same direction - but the authors don't see this.

  - Authors don't seem convinced about the idea of an olfactory map

  - Authors also rule out the idea of pure path integration because they show that many bats returned to a different cave than the one they started out the evening from

  –

1.3.1 Comments

- Fig 2E: why would you use the p-value to show the *absence* of an effect? The p-value per se is hinged on so many other factors (eg.power, effect size, sample size), why not just report the raw data

-

Chapter 2

Harten et al. 2020, Science

The ontogeny of a mammalian cognitive map in the real world (Harten et al., 2020)

- *notes taken on 2020-07-16*

2.1 Introduction, Methods, Results

- Whether animals navigate using 'maps'or not remains a question. The ability to take shortcuts, or direct routes between two points is a hallmark of map based navigation.

- The main problem with studying animal namvigation in the wild is that we can never be sure that the animal has not taken an apparent 'shortcut' before.

- authors were able to GPS track 22 young Egyptian Fruit Bats (*Rousettus aegeyptiacus*) from their first flight out of the roost

- Young bats increased their home range over the course of ~70 nights, by which they had the same home range size as an adult.

- Individuals showed two types of broad flight behaviour over a night, 'exploratory', where they explored for new trees, and nights where they visited prviously visted trees.

- Evidence to support the fact that the shortcuts were intentional:
  - shortcut were as straight as familiar routes ('commutes')
  - individuals seem to head in the direction of their target from the start of the 'shortcut'

- – the 'shortcuts' could not be replicated by a random correlated walk (*this seems like a bit of a straw man null model (link), especially since the data is clearly so directional. The authors also specifically mention 'but without any navigational goal'*)

- *Bats performed both shortcuts and long-cuts from their first day outside,* - this is pretty impressive, but this also makes me think that the bats may actually be relying on a kind of path integration. Is their apatial memory so good that they can start mapping things over the course of one night? Is it possible that the bats may actually be using a beacon-type stategy to find their way around?

- Authors rule out olfaction and sound based cues by comparing wind direction and actual recordings made on the backpack tags. Both don't show support.

- *'Bats that were closer to the translocation release point before the translocation night did not necessarily navigate home better, once again contradicting the template-matching hypothesis'* , the authors also go on to follow and say that bats that flew higher were better able to find their way back. This actually doesn't rule out the template matching hypothesis either, because it might just mean that bats that flew higher had larger access to the area below, to form a 'higher SNR' template perhaps...

- The authors do also admit that the navigation behaviou they observed may be a result of multiple navigation strategies: *'..., navigation is a complex behavior that probably does not always rely on a single strategy'*

-

2.1.1   Comments

- 'How animals navigate over large-scale environments remains a riddle', first line of the abstract starts with a rather bold statement. Is this statement really true for all animals, am under the impression that there is a large body of work for at least some animals.

- '*We documented how young pupsdeveloped their visual-based map*' - interesting, does this mean, echolocation develops later, or that the bats are known to use primarily vision for their navigation?

- Remember listening to a talk by Lee Harten in Konstanz ASAB, where she also presented work on the flight behaviour of the mother, who carried her pups around, and how the mother used to leave the pup in one tree, and so on. Do the authors discuss the implications of this type of memory on the shortcut taking ability of the young bats? *Yes, the authors have taken care of this, in the SI, they clearly state the mother and the pups*

were brought into an indoor facility, and the pups were kept indoors until they could fly

- *What about bats flying together?*, or encountering each other? Their in-hourse colony data kind of excludes this idea because the individual bats arrive alone, and are spaced b a few minutes. This is not the most rigorous evidence, but is still pretty indicatibe, eg. even in Orlova Chuka (and other caves), you can see the bats arrive alone in the morning. Howeer, this still doesn't really exclude the fact that bats may be encountering each other at some point over the course of the night.

-

Chapter 3

Wikelski et al. 2020

Potential short-term eathquake forecasting by farm animal monitoring (Wikelski et al., 2020)

- weird animal behaviour just before earthquakes habe been reported, including dramatic cases where snakes and rats came out of their winter burrows during the winter in the 1975 Haicheng earthquake

- finding reliable changes in animal behaviour is tough because animal the animal behaviour needs to be monitored before and during the earthquake.

- Authors were able to overcome some of the limitations in the data this time by tagging multiple farm animals with high-resolution GPS tags that were equipped with many types of sensors

- Authors measured the behaviour of animals at the M6.6 Norcia earthquake that happened on 2016

3.1   Methods

- Animals chosen from a farm based on which ones the animals thought were most sensitive to the earthquakes.

- Two tagging periods, once before + during the earthquake, once after

3.2   Data description

- *Between ..., the animals experienced a total of 5,304 earthquakes with M > 0.4 ...and from ... a total of 12,948 ...* didn't realise that earthquakes were so frequent in some areas.

- The 'hypocenters' of the earthquakes were anywhere between 5-28 km from the farm *–all relatively close by!*

3.2.1   Results

- Find a negative correlation between time of increased animal activity and earthquake intensity. For earthquakes ≥ 4 M, the animals responded earlier to quakes that were closer to the farm, and later to those that were further away from the farm.

- "Warning times" ranged from 1-15 hours

- Animals seemed to be more sensitive to earthquakes in closed buildings - but there may be a seasonal factor in the observations too

- What are the possible cues the animals are using to detect/respond to these earthquakes

  - The inverse relation hints at a diffusive type process. "air ionization at pressurised rock surfaces" – diffusing into the air, to which the animals may be responding to.

3.2.2   Overall thoughts

- very intersting paper, which quantifies something which has been known but has now been studied in greater detail through this new technology.

- authors also suggest a future experimental setup where a series of animal moinorting stations could be used to predict the position and time of arrival of an earthquake.

Chapter 4

Ratcliffe et al. 2004, Can. J. Zool.

Conspecifics influence call design in the Brazilian free-tailed bat, Tadarida brasiliensis (Ratcliffe et al., 2004)

- *notes taken on 2020-07-21*

4.1 Introduction

- echolocation is pretty flexible and the emitted calls vary a lot based on the type of prey being caught, and the presence of conspecifics
- This paper is a kind of offspring of another (Avila-Flores 2003), where the authors saw that there was more call variation when bats flew together than when alone.
- Authors try to estimate this variation by comparing observed pairs of bat calls and 'virtual' pairs of bat calls.

4.2 Methods

- Free flying *Tadarida brasiliensis* recorded in three locations in Mexico City, of bats flying over a lake, and two others in open areas (park and city square)
- 30 sequences from each location obtained (15 single bat passes + 15 paired passes).
- Authors specifically chose call sequences with no overlap of bat calls.

17

4.3 Analysis

- To compare the single vs paired call behaviour, the difference in the mean values was used
- The difference in the mean, Δ_{mean} was used to compare if a bat showed alterations to its echolocation when alone vs. when with another bats.
- To understand if bats actually altered their call parameters when flying in groups the authors calculated the pair difference $bat1_{measurement} - bat2_{measurement}$ for actual observed pairs of bat call sequences, and those of virtual pairs.
- The pair difference was calculated by subtracting the means of the $bat1_{measurement}$ and $bat2_{measurement}$

4.4 Results

- Authors didn't find any statistical difference in call parameters across locations and so decided to pool them all together (and thus used single bat call sequences from multiple locations while making virtual pairs). (*See comments*)
- Authors found no inter-individual call differences across observed and virtual pairs for the parameters 1) call duration 2) bandwidth , and found differences in the peak frequency. (*See comments*)
- The authors also saw social calls in paired audio files.
- The authors are indeed careful while ending '*To be even-handed….changes in peak frequency which we found…are neither jamming avoidance nor air traffic control, but serve another and as yet undetermined communicative function*'

4.4.1 Comments

- authors state '*Although referred to as jamming avoidance.., support for this interpretation is not as strong as that for the jamming avoidance response in electric fish'*. They also go on to state that labortatory studies seem to have 'met with some success'. In general, does this support the fact that bats in the field may actually not be showing dramatic changes in calls at all - it's only when they're put into a quiet unusual lab context that they begin to show changes. Perhaps, this strong response is seen because the animals have gotten used to flying under quiest conditions, and are now suddenly challenged?

- how do the authors actually know that the two bats in a paired call sequence were flying together/in close proximity? '*We assinged each recorded seequence to one of two situations…..and two bats flying close proximity*'

- this is a rather vague definition. With a single bat detector, it is not possible to track the bats in 3d, at most one can check if the waveforms of the two bat calls are similar, but here too it only means that the two bats were flying in the same radius from the bat detector. It is somewhat crude, though admittedly, the best possible criteria given the current instrument.

- Authors actually pooled call sequences from multiple locations and made virtual call pairs from this pool. This will actually have the effect of *increasing* the diversity of the observed data? Also, it is known from tracking studies that individual bats tend of have favourit foraging sites - and so in some sense, it is expected that there will be a 'local' flavour to the data. It would have been nice to see the authors perform the same analysis without pooling across locations.

- The authors find differences in peak frequency between real and virtual pairs. This could really be an effect of where the bats were flying while in pairs vs when they were alone. The simplest explanation is that the bats may be flying further away from the microphone, thus leading to different SNRs at recording - which then leads to different peak frequencies. The difference in the peak frequency could really be an artifact and not a real effect. Moreover, given the authors state that they do see social calls, the line between social call and echolocation call is a thin one, which means, perhaps the authors are seeing this effect? The authors state this themselves *'Our recordings suggest a continuum in call features between echolocation and social calls…'*

-

Chapter 5

Benediktova et al. 2020

Magnetic alignment enhances homing efficiency of hunting dogs (Benediktová et al., 2020)

- *Notes taken on 2020-07-22*

- 'homing' behaviour has been shown in many animals, and could be driven by multiple cues. Homing behaviour in non-migratory species still needs to be excplored more.

- Reports of dogs that found their way back even when displaced beyond their usual home range, and without access to familiar visual cues.

- Hunting dogs have been bred by humans to chase game and then return to the start point if not followed by the hunter. These 'hunts' can ve a few hundreds to thousands of meter long.

- These dogs may find their way back either using their own scent trail - which the authors called 'tracking' or, by actually navigating using the landmarks or information gained during the onward travel, which the authors call 'scouting'

5.1   Methods

- The hunting dogs were left to innately follow the olfactory tracks of game animals in the wildlife reserve
- Small dogs were used, and so there was no 'physical threat' to wild animals
- All dogs were tagged with GPS collars, and some of the trials with dogs also had a camera involved.
- Trials were performed with single dogs, and in areas free of high voltage power lines, roads or buildings and across the day and season (across the

year)
- There needs to be some more clarity in terms of when a trial started (see *Comments*)
- Owners hid behind a tree to reduce visual beaconing
- Only excursions >200 metres were considered for this study
- Authors split all excursions into 10 equal parts, and assigned phases to them based on the average speed of the part.

5.1.1   Return strategies

- Dog return strategies could be categorised into two classes
  - tracking: the inbound and outbound trajectory are the same , and the inbound-outbound trajectories were <30 m apart
  - scouting: a new woute was taken, and the inbound-outbound trajectories were >30 m apart

5.2   Results

- in ~59% of all trials, dogs ended up tracking their way back, while in 33% of the trials, dogs ended up scouting their way back, and in 8% of the trials, dogs used a mic of scouting and tracking
- Scouting dogs showed a higher average speed while returning (because they tended to take shorter routes)
- Trials bbegan and ended with no particular bias in the compass directions, though trials which were in the north-south axis, showed more efficient homing runs
- wind direction and sun location didn't seem to make a difference
- dogs returning with a scouting strategy seemed to begin by first travelling along the north-south axis, irrespective of the actual location of the destination

5.3   Discussion

- Authors convincingly rule out the possibility of multiple cues begin used(vision, olfaction, celestial cues). Thedogs are too short to see far, and the forest is pretty opaque. Olfactory cues are likely to change a lot, though one can't rule it out completely
- The sun's location/polarised light – this they haven't been fully able to elimination - but they indicate that there were trials where the sun wa blocked by clouds - which would reduce the strength of polarised light 'map'

- Authors do highlight the fact that path integration is definitely one possible mechanism by which the dogs in this tudy may be finding their way back, but the question of the north-south run still remains - and it may in fact serve as a kind of recalibration run, over which the errors accumulated over path integration can somehow be 'corrected'?

5.4 Overall thoughts / Comments

- Really smoth introduction - easy to follow
- It's not entirely clear how the trial exactly proceeded. The dogs were brought to a site, and then left to roam free.
    - *The handheld GPS device was programmed to indicate when the dog had travelled 100 m from the position of the owner. At this moment (designated as 'excursion start') the owners stopped walking...*. This is the confusing part, because, what was the owner doing before that? Were the dog and the owner walking together (unleashed dog) on a trail, and then when the dog roamed far away enough, the trial was considered started. But this still elaves some room for error in terms, of wehere the dog left the 'main trail' and where the owner was when the trial started.....confusing
- In figure 4, bottom plot - authors plot the *log* of the inbound track length against the raw beeline difference. This is a bit misleading perhaps? In general, yes, the log-transform will tend to show lesser variation - but in general, there will always be a linear correlation between te distance the animal travelled and the beeline distance ...especially if there is some kind of navigation in place... - am I being over-cautious here or not - can't make out.
- Pretty cool study, this study ads to the general observation of so many magnetically related behaviours (dogs peeing/pooping facing one direction (where did i read this), arctic foxes always facing one stereotyped direction before ponching onto their prey underneath the snow, and cows facing in one direction while grazing)

Chapter 6

Habersetzer 1981

Adaptive echolocation sounds in the bat Rhinopoma hardwickei (Habersetzer, 1981)

- *Notes taken on 2020-07-22, 23*

- *Rhinopoma hardwickei* behaviour studied as they emerge from their caves.

6.1   Methods

- Speed of the bat flight measured with a stroboscipic camera
- Echolocation calls recorded with a B & K microphone placed at the cave entrance.
- *R. hardwickei* lives relatively close to the cave entrance, and are found with other batr species too. The bats flew at an altidtude of around 10-15metres above the ground
- Bats began to return to their roosts around 3:00-3:30 in the morning, but most of them returned later between 4:30-5:45 am.

6.2   Results

- The author showed (I guess for the first time then) that *R. hardwickei* can actually modulate its echolocation calls, ie. by showing that the calls change over the course of a landing from classical quasi-CF to very FM

- During emergence, bats only emitted FM calls of ~3ms duration

- Emerging clusters seemed to have a preferred direction, while single bats that flew out didn't seem to have any preferred flight direction

- Single bats also seemed to emit CF type calls, in contrast to bats that were flying in groups

- Groups of bats sometimes formed over foraging grounds and in the early morning at the cave entrance itself. The bats in these groups emitted CF type calls.

- Individual bats in these groups seemed to emit calls with unique CF frequencies (Fig. 4 bottom), in contrast to single bats which seem to emit at a common frequency band while flying alone (Fig. 5)

- The bats seem to split thei CF frequencies into 3 bands, 30, 32.5 and 35 kHz.

- **R. hardwickeii** Cf calls don't seem to have an iFM or tFM unlike rhinolophids. The author seuggests this might mean that these bats are getting time delay purely through the envelpe of the call itself. I'm also fascinated by how the laryngeal mechanism must differ between the two species. Here, the absence of a FM to start/end means the bat must release the pressure uniformly over a very short period of time.

-

6.3   Overall thoughts / Comments

- Pye (1972) also found something similar to this paper apparently - need to check it out.

Chapter 7

Pye 1972, J. Zool. Lond

Bimodal distribution of constant frequencies in some hipposiderid bats (Mammalia: Hipposideridae) (Pye, 1972)

- *notes taken on 2020-07-23* (incomplete notes, left the paper halfway through)

7.1 Introduction

- Bats of one species seem to show echolocation in a 'common frequency band', ie. all measured bats calling within a 3kHz bandwidth

- Schnitzler 1968 found a max of 2.5kHz deviation across *R. ferrumequinum* and *R. euryale*

- Three species recorded in a cave in Kenya were however found to show 'sub-bands' with no individuals calling in between two main call frequencies.

-

7.2 Methods

- CF peak frequency determined through Lissajous figures (see Comments)

7.3 Results

- *Hipposideros commersoni*:

27

- – Bat seemed to emit somewhat higher frequency than expected from its size
- – Individuals either emitted at 56 or 66 kHz
- – Also measuremed the inter-nostril difference, as this is known to be ~0.5X the emitted wavelength - the nostril lengths were indeed different

- *Triaenops afer*
  - – This species also showed a bimodal distribution of CF at either 79 or 88 kHz
  - – *Here it gets interesting!* When individual animal call traces were placed over each other, then the calls seemed to be somewhat continuously spaced. But when a group of bats were flown together, bats seems to emit calls clearly in one of two sub-bands
  - – Fig. 2 – could the difference in the 'combined frequency profile' vs individual frequencies just be a function of flight speed. Maybe when the bats are palced together in a cage, they tend to overall fly a bit faster/slower – leading to this difference in the emitted frequencies by individuals?
  - –

7.3.1   Comments

- interesting to see how tricky it was back then to determine the peak CF frequency, something we do so routinely with a quick FFT nowadays. Pye had to visualise Lissajous figures, (comparing the CF call recordings with signals of known frequency)

-

Chapter 8

Jones & Ransome 1993, Proc. R. Soc. Lond. B

Echolocation calls of bats are influenced by maternal effects and change over a lifetime (Jones and Ransome, 1993)

- *notes taken on 2020-07-23*

8.1 Introduction

- *R. ferrumequinum* emits CF calls of ~50ms long at 83kHz.

8.2 Methods

- Ringed wild bats were captured and recorded as they echolocated. Overall, 386 bats were recorded in the hand.
- Bats ranged from 1-28 years old

8.3 Results

- Age seemed increase the resting frequency (RF), with a high increase from yaer 1 to 2, and then a gradual increase.
- There was also a seasonal effect - where the lowest RFs were recorded during the winter, and highest values in summer.
- The RF of infant bats were correlated with the RF of the mother (See comments)

- The authors also find a correlation with the age of the mother and the sex of the infant on its RF. Pups born to younger mothers had higher RFs, than those born to older mothers

8.4   Discussion

- authors point out that changes in RF with season may have to do with body temperature - as even active bats may not be as warm as they are in the summer.
- The correlation between mother and pup RFs may be explained by some form of learning perhaps
- changes in RF over age could be the result of hearing loss (neural/cochlear)

8.4.1   Comments

- the x and y axed of Fig. 3 aren't equal make it hard to judge visually how strong the correlation is. Numbers-wise at least, the r value isn't particularly high.
- back in 1993, the processing times were quick! The paper was received 1 Feb. 1993, and accepted on 16 Feb. 1993!!

Chapter 9

Schuchmann & Siemers 2010, *PLoS One*

Variability in Echolocation Call Intensity in a Community of Horseshoe Bats: A Role for Resource Partitioning or Communication? (Schuchmann and Siemers, 2010)

- *notes taken on 2020-07-23,24*

9.1 Introduction

- Authors measure the source levels of a whole community of horseshoe bats near and around the Tabachka field station. They record call levels of 5 species.

- The CF distributions of the four-five species that geographically overlap in Europe, also overlap spectrally

- *Rhinolophus mehelyi* can tell apart conspecific calls in the midst of other overlapping species (See comments)

- Do some bats (that are calling at higher than allometrically predicted frequencies) evolve towards a 'private frequency' channel?

- Authors look at whether *R. mehelyi* produce louder calls to compensate for decrease in detection distance because they call at a higher call frequency for their given body size (Comments 2)

- The questions the authors ask are:
  - Are the call intensitites different across species, and could call intensities play a role in nice differentiation (Comments 3)

- Does *R. mehelyi*, the higher-than-allometrically predicted species, call at 'especially high' call intensities? (Comments 4)
- How does max. call intensity vary with body size, sex – and could it be used as an honest signal bats can use to identify conspecifics (Comments 5)

9.2   Methods

9.2.1   Study animals

- *R. mehelyi, euryale, ferrumequinum, hipposideros* caught from the area around the field station, *R. blasii* caught from the Eastern Rhodopes.
- Bats recorded in a large-ish room (8x4x2.5m) (Comments 4), and can thus be expected to emit louder calls than when in a cluttered environment

9.2.2   Call intensitites and acoustic analysis

- Done with a B & K mic placed 1m away fromt he bat's head - *this is a pretty neat and direct way to do it because this is a perching bat anyways!!* ... though, the clutter caused by the mic being right in front of the bat ...hmmm...

- There;s a large variation in the frequency resolution across the years. It also is a bit unclear whether the authors used an automatic or manual method – though my current reading seems to suggest a semi-manual method to estimate the peak frequency of a call

- To get call intensity, the authors seem to do a lot of call processing.

  - FIR bandpass filter , with a 128 order bandpass filter(Comments 6) and then proceeding to take the FFTs within the call....why do this when you can take the waveform value directly??)

-

9.3   Results

9.3.1   Overlap of frequency bands

- species showed overlap in their CF frequencies
  - *R. mehelyi, euryale, hipposideros* overlapped, while those of *R. ferrumequinum* and *R. hipposideros* were different

- Statistically, the three overlapping species differed (but the sample sizes weren't too big either..)

9.3.2 Call parameters

- No observed correlation between call intensity and call frequency
- *R. euryale* seemed to call 10-17 dB lower than the other species.
- Intra-individual variation of 0-334 Hz across species
- Intra-individual call intensity variation also of between 2.3 to 5.5 dB

9.4 Discussion

- authors suggest call intensity can be an informative cue that can be used to assess species identity. They propose that perching bats may be able to reconstruct the position and 'head-aim' of the other bat, and thus assess its source-level. *While it may be informative to tell apart* R. mehelyi*, which calls fainter than others - there may actually be other cues eg. FM sweeps, or even more 'voice' type characteristics embedded in the whole call that could be picked up by bats themselves?)

9.4.1 Comments

- 1. Need to read ref. 24, Schuchmann & Siemers 2010 (Am. Nat.) - a habituation-dishabituation study

- 2. How does the *allometry* matter per se. The only thing that will matter for the detection distance is the source level and the emitted frequency (unclear)

- 3. It seems a bit tenuous, and obvious- but it's also hard to show a connection well?. Yes, low source level bats may only be able to detect insects from nearby, and high SL bats can detect insects from far away – not sure, but it sounds vague

- 4. I don't necessarily see a reason for it to call at higher source levels than what other species with similar CF do... not sure, why it would call at 'especially high' call intensities.

- 5. Probably this is the 'big flight room' at Tabachka?

- 6. Vaguely remember how using a very high order filter can mess with the signal structure itself (altering phases etc.). If the signal structure itself is altered so much, then what is the guarantee that the peak eq. source level obtained in this study is valid? If anything at all, the call intensites reported in this study may actually be at the lower

end of what is known because of the heavy filtering that has been done?

- 7. Odd grammar and typos in a bunch of places.

Chapter 10

Halsey 2019, *Biology Letters*

The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? (Halsey, 2019)

- *notes taken on 2020-07-24*

10.1 Introduction

- The p value has been used and abused, and in some sense has contributed to the reproducibility crisis in science

- The p value is based on the null hypothesis being true, and therefore doesn;t actually provide any information on how true the alternate hypothesis is – which is actually our question.

- If the p-value is 'high'/non-significant – *it doesn't tell us anything because there is an open question if our method could pick up a difference given the current sample size?

- '***Moreover, with a big enough sample size, inevitably the null hypothesis will be rejected; perversely, a p-value based statistical result is as informative about our sampole as it is about our hypothesis***' (See comments)

- p-values can vary between replicates of the same study even when statistical power is high! (See comments)

- interpreting p-values as signficant or not (dichotomously) encourages failed experiment replication — hmmm

- Even with 80% power —> prob. of getting two studies/replicates significant is 0.64 (0.8^2), while getting even one of these two studies/replicates significant is $0.32(2 \times 0.8 \times 0.2)$

- '*th ep-value is typically highly imprecise about theamount of evidence against the null hypothesis, and thuspshould be considered as providing only loose, first passevidence about the phenomenon being studied*'

10.2   Alternates to p-value testing:

10.2.1   P-value prediction interval

- the p-value is attractive, and a nice objective indicator of evidence against the null hypothesis
- author suggests presenting *variability* of p-values across samples, because p-values can be so variable : the 'p-value prediction interval' Cumming 2008

10.2.2   Estimate likelihood of false positive

- This method requires a prior 'feeling'/idea of how likely it is that there will be an effect , this is kind of hard to come by for observational studies, or those where there's no literature to say anything ??
- The author also provides an 'inverse' option, where the inputs are the obtained p-value + the power of the test. If the false positive risk is set at 5% ($\alpha = 0.05$), then the tool will provide an estimate of your prior expectation that the treatment will have an effect

10.3   Effect sizes and confidence intervals

- The effect size (differnce in mean between the groups) is easy to estimate - and a 95% confidence interval can be reported along with it.
- The confidence interval based approach with full reporting of the info prevents false claims from being made (yes/no effect)

10.4 Bayes factor: comparitive evidence of the null and alternate hypotheses

- the controversy about the 'subjectivity' in defining the prior can be solved by running a sensitivity analysis that explores a range of prior definitions
- the 'simplified Bayes factor' (SIB) is an alternate index which can be used to assess (Seems to be more relevant for planned studies and their stopping rules, also see Comments)

10.5 Akaike information criterion

- The AIC provides an idea of how well your model represents reality - and handles the tradeoff berween model complexity and over-fitting
- AIC is nice to compare between models, but it doesn't provide an absolute estimate of which model is the best. Among the models that are under consideration, it may very well be the case that the one with the lowest AIC doesn't capture the variation in the data as well as the other models - a check of the model's actual fit still needs to be done.

10.6 Conclusion

- '***Primarily, the argument goes, you should prioritize interpretation of graphical plots of your data, where possible, and treat statistical analyses as supporting or confirmatory information***'

10.6.1 Comments

- A low p-value can thus be a result of a high effect size for a given sample size, or a very large sample size for a given small effect size. So, how does one go about disentangling this issue then?? Perhaps with power analyses – which then adds to the amount of work that needs to be done anyway..?
- the p-value is not stable...hmm, even across replicates - this is kind of scary to know. ie. in (Halsey et al., 2015), they state *'The reason for this is that unless statistical power is very high, the P value exhibits wide sample-to-sample variability and thus does not reliably indicate the strength of evidence against the null hypothesis'*, and *'If statistical power is limited, regardless of whether the P value returned from a statistical test is low or high, a repeat of the same experiment will likely result in a substantially different P value'*
- Not sure how different this simplified Bayes factor is from a p-value - I also don't understand how it overcomes the problems associated with

dichotomous type p-value testing. Moreover, if the p-value itself is 'fickle', then it will also result in unstable SIB estimates too – how to overcome this problem...

- haha, nice - the author also provides a link to another paper Sneddon, Lewis, Bury 2017, J. Exp. Biol., where Box 2 has a template passage highlighting the problems of p-values!

Chapter 11

Halsey, Curran-Everett, Vowler, Drummond, 2015 *Nature Methods*

The fickle P value generates irreproducible results (Halsey et al., 2015)

- *notes taken on 2020-07-24*

11.1 Introduction

- p-values often used without understanding the statistical power in play
- statistical power is often only considered after an 'effect' could not be detected
- unless statistical power is very high, p-values vary a lot from sample to sample
- Ronald Fisher actually intended p-values to be used as a continuous indication of support for the null hypothesis
    - but, here too - if the statistical power is already low, no matter what - the p-value will not be reliable!!

11.2 statistical power is generally low −> p values can't be trusted

- Even when statistical power is close to 90%, p-values are not stable

- Considering that many studies are designed with 80% power, this means there is probably even more variation in p-values to be expected
- low statistical power ——> low sample sizes –> higher data variability

11.3   Exaggerated effect sizes

- A low-power test will give a low p-value only when the samples are very unrepresentative (ie. come from the edges of the distribution) – but it can definitely happen (Comments)

11.3.1   Comments

- the possibility of extreme samples contributing to low p-values, may also lead to a dramatic effect size. In some sense, there may be no *statistical* way to overcome this problem - the only way to resolve this situation may be to increase sample size?

-

Chapter 12

Li et al. 2014, *Behav Ecol Sociobiol* (Incomplete)

Behavioral responses to echolocation calls from sympatric heterospecific bats: implication for interspecific competition (Li et al., 2014)

- *notes taken on 2020-07-27*

12.1   Introduction

- sympatric species with simlar diets and ecologies could share information and gain mutual benefit through this process. Mutual recognition is required for mutual benefit to occur.
- agression between species at the same trWophic level should be selected against by natural selection as it increases the risk of predation
- In free-flying bats, species recognition has been shown to be related to simlarity in call structure and diet ecologies
- *'Interspecific recognition may thrterefore be necessary to adjust the strength of interspecific interactions'* - what is meant by this line - adjusted by whom?
- the trophic niche index - how 'linear'/uniform is it as an index?
- Authors tested which bats responded to which bats, and checked if call similarity and trophic overlap explained the observed correlation
- 3 species were tested, among the 4 species that are commonly found together:
  - *Rhinolophus macrotis*
  - *R. macrotis*
  - *R. lepidus* All the rhinolophid species were observed in the expeirments –but there were no *Asellia*'s in this experiment. (*Asellia*

41

stoliczkanus)

(Comments 1)

12.2 Methods

-

12.3 Results

12.4 Discussion

12.4.1 Comments

Chapter 13

Ming et al. 2020, *PNAS*

How frequency hopping suppresses pulse-echo ambiguity in bat biosonar (Ming et al., 2020)

- *notes taken on 2020-07-28*

13.1   Introduction

- In cluttered environments, echoes that arrive after the next call can cause call-echo ambiguity.
- When interpulse interval (IPI) is > 'echo-epoch' (time taken for the last echoes in the acoustic scene to arrive, α 'depth' of the scene) - then there is no ambiguity
- Echoes from the previous call that arrive can appear to be objects that are very close by.
- However, in cluttered conditions, bats reduce their IPI to keep up the update rate, which means the chance of such ambiguities is higher.
- 'Frequency hopping' helps to overcome this call-echo matching problem.
- *Eptesicus fuscus* shows shifts of upto 5-7 kHz between one call to the next while doing this kind of echo hopping. However, even a shift of 5-7 kHz means an overlap of 70-80% from one call to the other (Comments 1)

13.1.1   Quick primer on echo delay estimation

- In manmade ranging systems (radar/sonar), all frequencies in a wideband sweep contribute equally to the matched filtering.

- In *E. fuscus*, the first harmonic is required for echoes need to have a first harmonic to be 'registered', but the second harmonic (FM2) plays a stronger role in modulating echo delay perception.
- Authors test the specific idea that even *within* the first harmonic itself, *E. fuscus* only needs a limited region of the FM1 to detect echoes.

13.2 Methods

- Authors test if 1) only the lowest paer of FM1 frequencies are required for delay perception 2) unclear what is being *tested* here..*'Second, the small, seemingly not very effective, size of the frequency hopping by just a few kilohertz ....suggests that the hypothesized necessity of these lowest FM1 frequencies may be the key to why the harmonics are processed asymmetrically'*

- More and more of the lower frequencies were filtered out and the bats were asked to detect the S+ (Comments 3)

- If the authors want to understand which parts of the echo are important - why do they focus on an AFC (alternate forced choice) experiment where the task is to tell apart two echoes - with different arrival delays *and* different structures (single glint vs double glint) (Comments 4)

- It seems like *both* phantom echoes were high/low passed filtered - but this isn't explicityly stated in the paper!

- 4 *E. fuscus* were trained in an alternate force choice experiment

-

13.3 Results

- the region between 29-30 kHz is important for the bats to tell apart S+ and S-
- Confusing *'Finally, when all frequencies are removed from S+ by low-pass filtering at 20 kHz, there is no two-glint S+ at all,..'* - if 'all' frequencies are removed - there is no echo at all in principle, fo
- *'..vast proportion of frequencies do not support perception of echo delay unless the critical frequencies of 25 to 30 kHz are present.'*
- authors also cite previous studies where when FM bats were faced with CF playbacks, they responded strongly to playbacks between 24-28 kHz, and didn't respond at all when other frequencies were played back.

13.4   Discussion

- The authors seem to discuss the implications of the last few kilohertz results into how frequency hopping works – they haven't actually shown this directly through the experiments. The data only says the last few kilohertz is important for echo delay detection.
- The results indicate that the last few kilohertz is important - but at the same time don't point to a mechanism which allows the bat to detect echoes which are frequency shifted - how would the bat register the fact that the 'terminal frequency' has moved up after it's shifted its echolocation? This is not really discussed in the paper directly ?..

13.5   the SCAT model

- the SCAT model is a bio-inspired model which aims to mimic how the auditory system of bats work.
- delay perception is done by parallel bandpass filters which detect the time between sounds with energy in their bandpass region. Thus, they are able to tell the delay between call and echo fairly accurately.
- The idea the authors propose is essentially that the lower part of each call is the one that triggers echo detection.
- Imagine one call with terminal frequency at 20 kHz, and the next call with the terminal frequency at 25 kHz. Even if an echo from the previous call were to arrive - it'd have much lower frequencies than that of the current - call and wouldn't be 'sent' into the processing chain. (Comments 5)

13.5.1   Comments

1. How is this overlap actually calculated - it can be done in multiple ways…including/excluding the second harmonics etc..? As the authors themselves mention later - a shift of 5-7 kHz in the first harmonic corresponds to a shift of 10-14 kHz in the second harmonic.
2. Fig. 2 is the first figure that is referred in the paper (and the one that brings the phenomenon to the reader directly) - but, Fig.1 (related to a more technical point) is presented first - could have been rearranged.
3. S+ is introduced without any kind of explanation, and the figure is the main reference for the paper. This is kind of understandable given the word limits, but it's also kind of unfortunate because the reader's flow is interrupted each time by having to go back and forth between sections of the paper.
4. I'm not able to understand why the authors trained the bats to distinguish two echoes that differ in two characteristics. Am wondering how this would affect what can actualy be said from the experiments. ie. the results could

also be narrowly interpreted to say, 'bats require the lowest parts of their FM1 to tell apart a double glint echo from a single glint echo' ... It would have been cleaner if the authors had actually compared the bats ability to tell apart two echoes with the same structure, but arriving at different delays – this would actually say something about what is required to detect an echo, rather than a double glint vs a single glint echo. The task is made *easier* by the introduction of such different properties *'..question is whether this easy task still is easy (sic) if the lowest fre quencies in FM1 are removed from the S+ echoes'*, but the point is that it may also muddle the interpretation?

5. There seems to be a disconnect between what was done and what is being discussed. What was shown is that the end frequencies are important for echo discrimination (the glint structure X different delays confuses things according to me). But what is being discussed is how the frequency hopping is reliant on the lower frequencies of each call - the authors haven't really 'shown' this experimentally - but suggest a model which may explain it. There's something that I can't put into words better - need to give this a second try again.

6. The authors don't show the actual double glint echo - which is somewhat odd too.

13.5.2 Additional comments from AFEG Paper primer meeting on 2020-08-04 (Leonie, Paula, Holger, Thejasvi)

(Leonie leading, and other group members commenting)

- As the authors state, when the low pass is at 20kHz -there should be no double glint echo (silence on one side), which means the bats only choose the double glint, instead of just reverting to the side with one echo. This seems kind of odd

- The bat could actually be using frequency based cues, rather than the delay

- Not so many experimental details - a supplementary information could have been included

- The results only say that the last few kilohertz is important for telling apart S- and S+ - it doesn't necessarily say whether this part plays a role in glint enumeration (single vs double) or delay acuity!

- Authors haven;t really shown that in general the last few kilohertz aer important between calls, but the authors here have only shown that 25-30 kHz is important.

-

Chapter 14

Lakens, Daniel, blog post from *'The 20% statistician'*

Why you don't need to adjust your alpha level for all tests you'll do in your lifetime (Lakens, 2016)

- *notes taken on 2020-08-04*

14.1   Introduction

- Take a case where people are randomly assigned into two groups, and two unrelated dependent variables are measured. The groups are compared. True positive of no effect on both the variables is: 0.95 X 0.95 = 0.9025. Probability of at least one of the variables showing up as significantly different is: 1 - 0.9025 = **0.0975** !! How to control such erroneous false positives?

- There are a bunch of methods which can control the 'error rates':
    1. Bonferroni correction
    2. Holm-Bonferroni sequential procedure

- When the # of statistical tests increase, it may be better to control false discovery rate, rather than error rates.

14.2   The Bonferroni correction

- *'The Bonferroni correction controls the familywise error rate..' what is family wise error rate - See* Comments

- *'..is not straight forward is that error control does not just aim to control the number of erroneous statistical inferences, but the number of erroneous theoretical inferences.'.* Is the author trying to say that corrections need to be done when a bunch of tests are trying to answer the same question, with a common dataset - don't fully understand

- Experiment wise Type I error rates (*I guess this is the famous p value?*) can be calculated from the setup of the stats? The example given by the author here is in a 2x2x2 ANOVA (See Comments 2) - there are seven tests that are done, and so when a 0.05 α is used each time, this make the probability of getting at least one effect at least 30%. *(Comments 3)*

- The author also gives a more experimental example. The case of comparing predictions from two theories. One theory predicts interaction in a 2x2 ANOVA, and the other doesn't predict an interaction, but at least one main effect. (A 2x2 ANOVA tests 3 null hypotheses, two main effects and one interaction).

- A naive way to approach the Bonferroni correction is to set the α to $\frac{\alpha}{3}$

- But here the author points out that because theoretically the two theories predict different outcomes, the α also needs to corrected separately. Since Theory B predicts at least one significant main effect (out of 2 main effects) - we can set the new α to $\alpha/2$ and so if the p $< \alpha/2$ we will accept Theory B

- Since Theory A predicts only one interaction being significant, we don't really need to correct for anything. (Comments 4)

- Maintaining a sensible α value with multiple testing corrections allows you to be wrong at most at the expected α rate.

- Should you correct for all the tests you will do over your lifetime? Yes, but only if all the data/work you've done is going into testing one single theory !

- *'..criticism on corrections for multiple comparisons is that it is strange that the conclusions a researcher draws depends on the number of additional tests a researcher performs.'* - here the authors cites the case of a 2 group comparison being significant at 0.05 (p=0.04), but when including a second dependent variable, the α is now 0.025, which means the first variable data no longer rejects the null hypothesis. *..'Lowering alpha levels is a mathematical necessity when you want to control error rates, but it is not needed if ..you..quantify relative likelihoods of the data under different hypotheses.'.* Does the author mean that the comparison of the two groups for each of the dependent variables corresopnds to two independent null hypotheses? In what way is this different from the 2x2x2 ANOVA that was just discussed before?

- You can also increase the alpha, and sometimes this makes sense to do so - ie. while pre-registering two studies, you can set the alpha at 0.2236, and decide that only if *both* studies show $p < 0.2236$ will the null hypothesis be rejected. This is effectively the same as setting the overall p at 0.05 because $0.2236 \times 0.2236 = 0.05$ - [Comments 5]

- *'There is only one reason to calculate p-values, and that is to control Type 1 error rates using a Neyman-Pearson approach.'* - the focus here is on controlling false positives.

14.3 Comments/Questions

1. what is a 'familywise error rate': The probability of making at least on Type I error (False Positive): $FWE \leq 1 - (1 - \alpha_{test})^{Number\ of\ comparisons}$. eg. at an α of 0.05 and with 10 tests, the FWE is : $\leq (1 - 0.05)^{10} = 0.401$ (thanks to this page)
2. 2x2x2 ANOVA is when there are three variables and two groups in each variable. ie. imagine you're testing a drugs efficacy, and you have three variables (Thanks to this page) : sex, dosage level and treatment/control. Within Sex there's male and female, within dosage level there's high and low, and then there's treatment/control. There are seven p-values because it performs a series of comparisons within the 2x2x2 complex:
   - 3 tests comparing the groups within each of the variables (eg. male vs female, treatment vs control, high vs low)
   - 3 tests testing two-way interactions: pool data from Sex, effect of treatment vs control in low and high dose groups
   - 1 test on three way interaction: *'there is no three way interaction among all three factors'*
3. I get this example, within one statistical test, there are seven comparisons being made - and when each of these have an alpha value, there can be a false positive - what about when you are performing single comparisons of multiple variables from two groups?
4. I guess one of the main issues that the author is trying to bring out is that by naively correcting and setting α values very low, we risk running into Type II (false negative) errors! By setting
5. What would have happened if the alpha was still set at 0.05 for both studies, then the effective false positive rate would have been 0.025 – which is perhaps too conservative? Also, with an alpha set at 0.2236, the probability of a false negative is:

Chapter 15

Culina et al. 2020, *PLoS Biology*

Low availability of code in ecology: A call for urgent action (Culina et al., 2020)

- *notes taken on 2020-08-11*

15.1   Introduction

- There is more and more awareness on the part of journals and authors to be more open about the data and code.
- FAIR principles (See this paper):
    - Findable
    - Accessible
    - Interoperable
    - Reusable

15.2   Methods

- Authors try to quantify how oftent scientists share their code, and were able to identify 346 non-molecular studies between 2015 and 2019
    - Has code availability increased over time?
    - How well are authors actually following prescribed practices related to code sharing (FAIR principles, etc.)
    - *'limits to computational reproducibility'* (Comments 1)

15.3 Results

- The number of journals requiring code sharing has gone up -> positive trend.

15.3.1 Low code availability

- 90% of all software *used* was either GUI or command line code, though the authors even went to the extent of including papers that had screenshots and instructions on how to use the program. (Comments 2)
- Only 27% of all surveyed articles had code that had some or all of their code available. (Comments 3)

15.3.2 Findability, accessibility, and reusability

- Authors point out a very relevant points '*To be found, code availability should ideally be stated within the article or, if not, within the supplementary material of the article.*'
- 24% of all articles that actually provided the code didn't state it at all - but instead had broad references to the supplementary infor.
- Authors point out that even though archiving code with the supplementary information of the journal works - journal access is not always universal - and this can affect accessibility!
- Authors suggest the use of Dryad and Zenodo to archive code - which makes sense.
- '*We oped to simply evaluate whether the published code had some form of documentation (e.g.README), either as an accompanying document or embedded at the beginning of the code, and whether the code had inline comments explaining parts of the code.*' (Comments 4)

15.3.3 Limits to computational reproducibility

- Only 21% of all articles provided both code and data together
- Previous estimates of computational reproducibility showed only between 18-80%.
- Even in the papers which provided both data + code $->$ 26% ofc the papers used proprietary software (Comments 5)

15.3.4 Comments

1. Not clear what is meant by 'limits to computational reproducibility' here.

2. Personally, I don't quite consider screenshot based records very repro-ducible, because they don't scale well with the number of steps. There are so many details to take care of often, and this is just a very cumber-some way to record and to read and try to reproduce an analysis.

3. Woah, even though the authors were relatively lax in their definition of what constitutes reproducible code – even then just about one in three surveyed papers had the associated code with it.

4. Here too the authors have actually been generous with their criterion for what documentation actually is. I understand that most people will not have the time or training to document their code to the point where it is completely self-contained - but this is also the problem.

5. Another reason to avoid using proprietary software - it reduces the rate at which others can use your code *and*, most importantly, excludes the quick replication of your results if someone wants to 'fact check' it. In general, I can see how certain proprietary softwares may do the job well, but it also means it's an ivory tower - and no one can look into it. However, what can be done is to ensure a coomon set of coding standards are being followed nonetheless (code documentation, tests, READMEs, etc). I'm sure there must be a bigger discussion about this in the literature, as these are my first guesses/thoughts.

6. The authors seemed to have used an automated tool to find their papers - 'Rayyan' - interesting. This tool may be helpful in the future perhaps?!

Chapter 16

Wilkinson et al. 2016, *Scientific Data*

The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al., 2016)

- *notes taken on 2020-08-11*

16.1 Introduction

- *'data stewardship includes the notion of 'long-tem care' digital assets, with the goal that they should be discovered and re-used for downstream investigations*
- *'Humans, however, are not the only critical stakeholders in the milieu of scientific data.'* - authors refer to the fact that 'computational agents' (like bots for eg?) are becoming equally important in the curation/searching of knowledge
- Many field-specific databases exist with their own formats and curating workflows - eg. protein databanks, astronomy
- There are more and more discipline agnostic databanks coming up nowadays - FigShare, Dryad, Mendeley DAta, EUDat, DANS, DataHub
- This increase in multiple databases and systems, may actually be making it harder to keep track of datasets for human and 'computational stakeholders'
- When a researcher is looking to compare the data they've generated with pre-existing datasets - how to go about this task?
  - the search needs to be passed through various filters, data formats, species etc. – it can get complicated finding the relevant result

- *'The goal is for scholarly digital objects of all kinds to become 'first class citizens' ....where the quality of the publication-and more importantly, the impact of the publication-is a function of its ability to be accurately and appropriately found, re-used, and cited over time, by all stakeholders, both human and mechanical.'*
- there are already machine actionable formats out there, eg. the ones used in many databanks and space sciences etc, but the aim is really to come up with something that is broad, so that a machine can decide on a dataset's relevance/contents even though it hasn't seen it before.
- This site https://fairsharing.org/ seems to be pretty useful in detailing many of the publicly accessible databases/data repositories, and also details the various file formats out there, among other things.

16.1.1   The significance of machines in data-rich research environments

- humans have a better understanding of what is relevant and what is not while searching, but human searching is not scalable.
- To enable automatic searches/curation - datasets need to be machine actionable', ie. have machine parsable identifiers, info on the type of objects, licenses, consent and etc.

16.1.2   Comments

- This paper talks about some ideas which are definitely going to be important.   One thing that confused me is that they talk about databases/repositories which are exemplary in the paper (DataVerse, FAIRDOM, ISA, Open PHACTS, etc), but don't mention any of the other ones (eg. Dryad, Zenodo, FigShare, etc) - a comparison would have been nice to understand why these were not included. I can also see that there is a clear discipline-based bias in where people upload their datasets based on where their colleagues do it too.
- In general, the comment about the growing number of databases/repositories is a very valid one. While reading this paper itself I began to realise how narrow my overall awareness for various databases is!

Chapter 17

Zhang et al. 2019, *J. Exp. Biol.*

Dynamic relationship between noseleaf and pinnae in echolocating hipposidered bats (Zhang et al., 2019)

- *notes taken on 2020-08-25*

17.1 Introduction

- Noseleafs in hipposiderids are known to shift in shape during call emission, and thus affect where the call is directed.
- Pinna movements are also thought to generate additional sensory information for the bat to use
- Authors test to see if noseleaf and pinna motions are coordinated during echolocation.
- Perform a canonical analysis to see if the pinna and noseleaf motions are somehow coordinated/occur together.
- *'We also expected that canonical correlation would detect a relationship between the motions. But a negative result of either analysis does not reject the hypothesis, because neither method is guaranteed to find any existing relationship. By demonstrating coordination between emission and reception dynamics, a positive result of this analysis could lay the foundation for a better understanding of active, dynamic sensing paradigms in bats and beyond.'* (Comments 1)

17.2 Methods

- *'We exposed the bats to a variety of changing biosonar targets to attract their attention and elicit exploratory biosonar behaviors triggered by novelty.'* : missing details! How were the bats placed, were they sitting, flying. Very sparse description, which doesn't give any details at all.

- A lot of detail given to the 3d camera tracking – which is not really required in my opinion. This is not really the novel contribution of this paper, and so these details can be reduced - and more focus on the actual science the authors performed

- Only a few dots numbered, the rest not -why? (Comments 2)

- *'The microphone was mounted at the same height as the platform on which the bat was positioned at a distance of 35 cm in front of the animal'* (Comments 3)

- Pinna movements classified into 'rigid' and 'non-rigid' (if all landmarks moved and were still $<=$ 2mm of their previous positions, the motion was called rigid)

- Closing and opening noseleaf motions occured with non-rigid pinna movements

17.3 Results

- K-means clustering separated noseleaf motions into 4 groups: closing, opening, random and no motion.
- All closing noseleaf motions coincided with call emission, but only 3/74 coincided with opening motions.
- *'closing and opening motions of the noseleaf were mostly accompanied by non-rigid pinna motions (68.95 to 90.41%)'* - what is the meaning the of 'to'. Should it be replaced by 'respectively', and if it is actually a min-max range, then what is the range calculated over.
- noseleaf and pinnae open and close together (Comments 4)

17.4 Discussion

- *'The qualitative and quantitative relationships demonstrated here are predicted by the hypothesis that the biosonar system of hipposiderid bats relies on the synergistic integration of emission and reception dynamics'* – where is this hypothesis explicitly stated?? The authors only state this in the introduction *'Our underlying hypothesis is that movements of noseleaf*

and pinnae are important to biosonar performance of rhinolophid and hipposiderid bats.'. The authors are claiming something that they did not show in their data, and the 'predictions' they are testing are not really predictions per se. - (Comments 5)

17.4.1 Comments

1. This is a somewhat odd statement because *any* statistical/analytical method cannot guarantee the detection of an effect. The ability of any stat/analytical method relies on so many things, the data, the sample size. On the other hand, even if a method 'detects' something, there is no guarantee that it is not a false positive - which has no biological meaning. This is a somewhat odd statement especially because it seems to only focus on the false negative side of the story, and overly favours a positive statistical/analytical result. Taken independently, this sentence bothers me somewhat.

2. In Fig 1C the bat has many more dots than are numbered. Why were these dots not used in the final analysis, or discussed?

3. Only now we find out that the bat was actually placed on a platform and the setup was placed in front of it.

4. Wonder what the sensory consequences of this are? What happens *after* the call? The echoes would typically return a few ms after the call - how fast are the pinnae responding to this, over what time-scales is the pinna motion actually happening.

5. In the introduction, the authors would like to test the hypothesis that noseleaf/pinna movements are important for biosonar performance. The prediction of noseleaf and pinna movements comes from nowhere - it is *not* a prediction, but they are *exploring* the connection!!! This is poor and misleading writing, which tries to fit a square into a triangle shaped hole. It's completely fine if the study is exploratory, but there is no need to sell it off as a study which comes from a theoretically based 'prediction'

Chapter 18

Goldshtein et al. 2020, *Current Biology*

Reinforcement Learning Enables Resource Partioning in Foraging Bats (Goldshtein et al., 2020)

- *notes taken on 2020-08-28*

18.1   Intro-Methods-Results

- '*Every evening, from late spring to mid-summer, tens of thousands of hungry lactating female lesser long-nosed bats (*Leptonycteris yerbabuenae*) emerge from their roost and navigate over the Sonoran Desert, seeking for nectar and pollen*' : Comment 2

- Something I'm seeing in papers and am responding to. 'We hypothesised.' without clearly stating the reason behind it.

- *..quality of a specific cactus can vary dramatically according to the number of flowers it opens on a given night* - somewhat misleading sentence. Comment 3

- Fig 1A. - showing the number of open flowers per cacti - could have shown a *discrete* colour bar. The use of a continuous color bar for integers is very confusing.

- '*Moreover, bats have to re-map the field every night.*' - not clear why, or what this statement is based on

- *Although we found a significant correlation between the number of flowers on consecutive nights, there was much inter-night variability* - the reference

to the figures aren't v clear : Comment 4, 5

- '*…because a conspecific might deplete it in the meanwhile (the bats' behaviour suggests they do not know whether a floewr was recently visited…)*' - S1E captioned erroneously perhaps Comments 6,7

- *'Finally, all of these data were used to develop a mathematical model..'*, hmmm, this is a bit semantic, but aren't evolutionary simulations , *simulations* and not 'mathematical models' as such.

- GPS tracked 17 bats, and then finally reconstructed the foraging sites of 8 bats. This must have been a lot of work over all! Perhaps, the authors even actually tagged more bats - and considering the problems with tag losses and animal recaptures - this must have been quite an effort.

- Commuting speed was 8.7 m/s ground speed (~31km/h), and commutes correspond to ~46% of the time spent away from the cave

- Even within a cactus field - each bat concentrated its foraging in a narrow part about 0.14 km^2 in area (0.14 km^2 is ~ 374x374 m^2)

- Plants that were visited $\leq$ 5 times were considered to belong to the bats' 'core cacti' - Comment 8

- *'Indeed, video analysis revealed that cacti with more flowers were visited significantly more often'* - this triggered the thought - do the authors actually report if bats visited multiple flowers per visit? This might be another way that bats learn to identify which cacti are worth visiting multiple times, and those which are not. Also did core cacti have more flowers than those around them?

- In the reinforcement learning simulations, the bat remembers the top ten cacti it visited according to the amount of nectar it provided at that point of time ('nectar weight'). The next cactus to be visited is determined based on its nectar weight and inversely proportional to its distance from the current plant.

- '*Importantly, we did not fit the data in order to find the learning rate α*' - this is not true. An evolutionary algorithm is one of many types of algorithms that can be used to estimate a parameter. All of these algorithms use some kind of minimisation/maximisation algorithm. An evolutionary algorithm just uses the fitness of the agent, which is in turn affected by the cost functions in the model itself. Moreover, the authors do use the actual data to parameterise the evolutionary simulations themselves. I don't see how the act of using an evolutionary algorithm means the parameter is not being fitted in the same way a common linear regression fits the data to get coefficients. This is logical wrong, and also misleading - adds an air of deeper sophistication than there actually is.

- Authors show that when compared to a trapline model or a random search model, the reinforcement model matched various aspects of the observed

dataset well (inter-visit distances, revisit time intervals)

- Authors also ran simulations with aggressive behaviours in the agents, unlimited cacti memory (instead of a fixed 10 plants) - and still saw the appearance of non-overlapping core plants across multiple individuals

- '*We prove analystically that...*' - missing reference to [29] (Korman et al. 2020, ArXiv preprint)

- '*Interestingly, a single-parameter model based on reinforcement learning was able to closely reproduce the bats' behavior*' - in some sense – is this really a single-parameter model? The learning rate α makes sense only in the context of the many other parameters that were used to setup the model right? Would the same learning rate be valid given that the bats flew at a different speed or the plants were placed at different spatial arrangements. This is not a fully generalisable result I feel....also isn't

-

18.1.1 Comments

1. The authors state *'Many species have been suggestd to follow a trapline, that is, to revisit the food sources in a repeating ordered manner...We thus hypothesized that lesser long-nosed bats would visit cacti in a sequenced manned.'*. Is this a legitimate form of hypothesising. Isn't it closer to expecting to *examining*/checking if the bats perform sequenced foraging? Am wondering if this kind of formulation is the effect of the NHST type thinking, where there is one hypothesis that needs to be formulated and either rejected or supported.
2. Like it as an opening line. Very graphic, and very evocative of what is being studied. The sentence could have been shorter though.
3. The sentence creates the impression that quality is somehow influenced by the number of flowers open - as if they were *two* variable. But instead, here, the authors actually define the *quality* of a cactus by the number of flowers it has open!
4. Comment about figure S1 in general. This is a very compact graph, that prevent leisurely inspection of the data. Annother approach would have been to give one page to each graph and legend. This is the supplementary information section after all. (Does Current Biology have page limits even for this section?)
5. Figures S1C,D: it's not clear to me why the 'independent model' looks the way it does. If there is no correlation at all between the last and current night there should be equal transition probability for any transition (0-10 flowers, 0-9, 1-10 etc..) - but here too there is some kind of clustering. Why is this the case - more details are required to explain the model. The explanation by itself is not detailed enough for criticism or replication.

6. S1E captioned 'Bats probably cannot identify which flowers *they* previously visted' - emphasis mine. - the caption refers to a bat knowing which flowers it previously visited while the main text refers to the idea that bats may not be able to tell if other conspecifics previously visited. Although in the following sentences the situation is corrected.

7. Figure S1F refers to the *order* of visitation, and not really to whether bats can tell if a flower has been previously visited or not. This is confusing. Also S1F1,2 are incredibly small figures

8. The authors choose the points at which the derivative of the probability mass function (PMF) drops to zero. I'm thinking about how generalised this approach is. It works and it intuitively makes sense for *this* particular PMF because the probability of increasing visits drops very drastically, and the distribution is right-skewed. The same definition would have failed if the PMF had been more flat, or bi-modal-ish. Another option would have been to use an X% cumulative probability approach, ie. the #of visits at which the X%ile of the data is seen. This is robust to the exact shape of the distribution.

9. What kind of 'cost function'/ metric was used. The exact parameterisation of this could have an effect of the types of decisions the agents end up showing.

Chapter 19

AUTHORS ET AL. YYYY, *J. STHING. STHING.*

BIG TITLE OUT HERE IN FULL (Schuchmann and Siemers, 2010)

- *notes taken on YYYY-MM-DD*

19.1 Introduction

19.2 Methods

19.3 Results

19.4 Discussion

19.4.1 Comments

Bibliography

Benediktová, K., Adámková, J., Svoboda, J., Painter, M. S., Bartoš, L., Nováková, P., Vynikalová, L., Hart, V., Phillips, J., and Burda, H. (2020). Magnetic alignment enhances homing efficiency of hunting dogs. *Elife*, 9:e55080.

Culina, A., van den Berg, I., Evans, S., and Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLoS biology*, 18(7):e3000763.

Goldshtein, A., Handel, M., Eitan, O., Bonstein, A., Shaler, T., Collet, S., Greif, S., Medellín, R. A., Emek, Y., Korman, A., et al. (2020). Reinforcement learning enables resource partitioning in foraging bats. *Current Biology*.

Habersetzer, J. (1981). Adaptive echolocation sounds in the batrhinopoma hardwickei. *Journal of Comparative Physiology*, 144(4):559–566.

Halsey, L. G. (2019). The reign of the <i>p</i>-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters*, 15(5):20190174.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., and Drummond, G. B. (2015). The fickle p value generates irreproducible results. *Nature methods*, 12(3):179–185.

Harten, L., Katz, A., Goldshtein, A., Handel, M., and Yovel, Y. (2020). The ontogeny of a mammalian cognitive map in the real world. *Science*, 369(6500):194–197.

Jones, G. and Ransome, R. D. (1993). Echolocation calls of bats are influenced by maternal effects and change over a lifetime. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 252(1334):125–128.

Lakens, D. (2016). Why you don't need to adjust your alpha level for all tests you'll do in your lifetime. Accessed: 2020-08-04.

Li, Y., Wang, J., Metzner, W., Luo, B., Jiang, T., Yang, S., Shi, L., Huang, X., Yue, X., and Feng, J. (2014). Behavioral responses to echolocation calls

from sympatric heterospecific bats: implications for interspecific competition. *Behavioral Ecology and Sociobiology*, 68(4):657–667.

Ming, C., Bates, M. E., and Simmons, J. A. (2020). How frequency hopping suppresses pulse-echo ambiguity in bat biosonar. *Proceedings of the National Academy of Sciences*, 117(29):17288–17295.

Pye, J. (1972). Bimodal distribution of constant frequencies in some hipposiderid bats (mammalia: Hipposideridae). *Journal of Zoology*, 166(3):323–335.

Ratcliffe, J. M., Hofstede, H. M. t., Avila-Flores, R., Fenton, M. B., McCracken, G. F., Biscardi, S., Blasko, J., Gillam, E., Orprecio, J., and Spanjer, G. (2004). Conspecifics influence call design in the brazilian free-tailed bat, tadarida brasiliensis. *Canadian Journal of Zoology*, 82(6):966–971.

Schuchmann, M. and Siemers, B. M. (2010). Variability in echolocation call intensity in a community of horseshoe bats: A role for resource partitioning or communication? *PLOS ONE*, 5(9):1–10.

Toledo, S., Shohami, D., Schiffner, I., Lourie, E., Orchan, Y., Bartan, Y., and Nathan, R. (2020). Cognitive map–based navigation in wild bats revealed by a new high-throughput tracking system. *Science*, 369(6500):188–193.

Wikelski, M., Mueller, U., Scocco, P., Catorci, A., Desinov, L. V., Belyaev, M. Y., Keim, D., Pohlmeier, W., Fechteler, G., and Martin Mai, P. (2020). Potential short-term earthquake forecasting by farm animal monitoring. *Ethology*, n/a(n/a).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9.

Zhang, S., Liu, Y., Tang, J., Ying, L., and Müller, R. (2019). Dynamic relationship between noseleaf and pinnae in echolocating hipposiderid bats. *Journal of Experimental Biology*, 222(18):jeb210252.