Title

Definition and interpretation effects: How different vigilance definitions can produce varied results

Authors

Andrew T. L. Allan[1,2*], Russell A. Hill[1,2,3]

Affiliations

1. Department of Anthropology, Durham University, Dawson Building, South Road, Durham DH1 3LE, UK.
2. Primate and Predator Project, Lajuma Research Centre, PO Box 522, Louis Trichardt 0920, South Africa.
3. Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa.

*Corresponding author.

Email: a.t.l.allan@durham.ac.uk

Phone: 01444 248873

Word count: 9690 (main text, including tables)

Abstract

Animals use vigilance to detect or monitor threats. While numerous aspects of vigilance have been studied across a wide range of species, little work has explored the methodological variation that has emerged across these studies. Different approaches in sampling designs, statistical analyses, and definitions can make cross-study comparisons challenging and potentially obscure our understanding of animal vigilance. In this study we explore two important components of vigilance definitions and ask: 1) whether definitions vary in their inter-observer agreement, and 2) whether using different definitions can create varied results within and across observers. Separate groups of 'experienced' and 'inexperienced' observers extracted data from video focal observations of wild chacma baboons, using four different definitions representative of the variation found within primate vigilance literature. In the first stage of analysis, we found that the four definitions varied in their inter-observer agreement, with only an operational looking definition performing well across both duration and frequency assessments, and an experienced/ inexperienced dichotomy. This suggests definitions vary in how well observers can converge on similar interpretations of the same definition. The second part of the analysis used the experienced group's data in a typical primate vigilance analysis and found results varied within observers across definitions, i.e., definition effects, and across observers within definitions, i.e., interpretation effects. Together these results suggest that variation in definitions and their interpretation could have a fundamental role in producing between-study differences in results. Future vigilance research must consider these factors and explore working towards a single framework for studying vigilance, particularly within taxonomic families. Without consistency, cross-study comparisons are likely to be challenging and future observational work on other behaviours may also benefit from exploring these types of definitional issues. For baboons, operationalised definitions appear the most consistent across observers, however, future research should explore its application in other taxa.

Highlights

- Variation in vigilance definitions has grown, both within and between species.
- Observers coded videos of baboons using four definitions derived from primate research.
- Definitions varied in their inter-observer agreement and the results produced.
- Operationalised definitions improved consistency of estimates across observers.
- Similar assessment frameworks are required in other taxa.

Keywords

Behaviour, consistency, definition, methods, reliability, vigilance

Introduction

Understanding how animals use vigilance to detect and avoid predators (amongst other threats) has proved a popular topic for animal behaviour research, with bird species generally receiving the most attention (Beauchamp, 2015). Birds have offered researchers an excellent study system, as their postural changes are relatively straightforward to monitor (Beauchamp, 2017), with 'head-raising' or 'head up' postures used as markers for vigilance (Beauchamp, 2018; Fernandez et al., 2003; Klett-mingo et al., 2016). Such definitions have been common in vigilance research, seemingly since the model of Pulliam (1973), which inferred 'head cocks' allowed birds to detect predators. But as vigilance research has grown, variation in vigilance definitions has become expansive, with a variety of definitions emerging (Allan & Hill, 2018; Beauchamp, 2015).

Potentially because of the predominant focus on vigilance patterns in birds, numerous mammalian vigilance studies have followed a similar postural paradigm in defining vigilance. For example, in a general review of mammalian vigilance patterns, Quenette (1990) defined vigilance as "a head lift interrupting the ongoing activity, and followed by a visual scanning of the environment". Comparable definitions have been used in studies on antelopes (Lian et al., 2007), sheep (Brown et al., 2010; Rieucau & Martin, 2008), kangaroos (Carter et al., 2009; Favreau et al., 2010), capybara (Yaber & Herrera, 1994), marmots (Ferrari et al., 2009; Shriner, 1998), squirrels (Arenz & Leger, 2000; Shriner, 1998), primates (Alberts, 1994), lizards (Ito & Mori, 2010; Javier & Perez-Mellado, 2000), and fish (Brandl & Bellwood, 2015). Nevertheless, as the taxonomic focus of vigilance work has broadened, variation in vigilance definitions has emerged. In most part this is due to different species having different postural capabilities and tendencies but can also be attributed to the specific focus of individual studies. Many studies on mammalian species have taken into account the distinction between quadrupedal and bipedal (Bednekoff & Blumstein, 2009; Blumstein et al., 2010; Quirici et al., 2008; Vasquez, 1997) or "vertical" postures (Unck et al., 2009). Vigilance research on birds has also applied additional constraints to vigilance, such as the requirement of "side to side movement of the head” (Elgar & Catterall, 1981), “stood still, with necks fully extended” (Burger & Gochfeld, 1988), or “extending (‘stretching’) the head upwards and looking around while standing straight” (Li et al., 2017). Similarly, in lizards, definitions have been based on a range of factors, including “head movement” when stationary (Ito & Mori, 2010), “eyes open” (Lanham & Bull, 2004), and “pause” in locomotor activity (Lopez & Martin, 2013), whilst Iberian wall lizard’s (*Podarcis hispanica*) were considered vigilant when immobile/paused and had their “head high, their eyes raised”, and “rarely moving their heads side to side” (Javier & Perez-Mellado, 2000).

Subtler variation can also be found across studies on similar species. For example the following definitions have been used for *Cebus* monkeys: “animal had its head-up and looked around, providing it was not inspecting vegetation or partners at close range” (van Shaik & van Noordwijk, 1989), “scanning intently at long range while alert and stationary” (Campos & Fedigan, 2014), “cautiously observing the surroundings, often with horizontal rotation of the head” (de Ruiter, 1986), and “visual inspection of surrounding area without a fixed gaze; turning head side to side (Fragaszy, 1990). Across a number of different geese species vigilance has typically been defined using “heap up” or “extreme head up” postures (Forslund, 1993; Kahlert, 2003; Shimada & Shimada, 2003); however, other examples include “head was above the level of its back” (Atkins et al., 2019), and “head held in an upright position, looking around and alert” (Tadeo & Gammell, 2018). The nuanced variation that has emerged in vigilance definitions may be required to sample the unique postural and behavioural traits of each species and may be ecologically valid and necessary from a methodological perspective. However, some variation appears to have emerged not out of necessity but through vigilance research lacking a consolidated framework. Little is known, however, about whether variation at the definition level could alter the distribution of datasets collected, and thus results, making comparisons of findings across different individuals, groups, and species challenging.

In addition, little work has explored the repeatability of individual definitions, i.e., how well different researchers or observers converge on similar interpretations of the same definitions. As the list of more technical definitions has expanded it is unclear is whether certain requirements are challenging for observers to reach agreement on and so lead to inconsistencies between independent studies. Some definitions contain elements that ask observers to interpret aspects of the animal's behaviours and state which could be prone to varied interpretations, e.g., scanning intently (Campos & Fedigan, 2014), visual scanning of the environment (Quenette, 1990), looking around and alert (Tadeo & Gammell, 2018); however, postural definitions such as head up or head raising could also vary in inter-observer agreement, e.g., do all researchers agree on the exact point the animal's head is up? Is this consistent across different studies and species?

In this study we use the variation in definitional ethos found throughout primate vigilance research as a framework to investigate these questions. Primates use a range of postures and have the ability to handle food items whilst looking around concurrently (Cowlishaw et al., 2004). Potentially as a result of these factors, primate vigilance studies have used a plethora of definitions despite being relatively understudied compared to other taxa in terms of vigilance (Beauchamp, 2015). This has potentially generated the highest degree of variability in vigilance definitions for a single taxon within the literature (see (Allan & Hill, 2018)). Nevertheless, there seems to be a clear dichotomy between studies interested in reactionary and those investigating preemptive aspects of vigilance. Studies focusing on reactionary vigilance appear to use postural changes associated with danger recognition such as "active visual search skyward with an exacerbated posture" (Boinski et al., 2003). In contrast, studies of pre-emptive vigilance focus on data collected in the absence of threatening events (Teichroeb, 2017). This is complicated in primates, however, as numerous species also need to monitor both within- and extra-group conspecific threats (see (Allan & Hill, 2018)). Pre-emptive vigilance markers are likely much more subtle (reducing the cost to the animal) and therefore challenging to identify (Allan & Hill, 2018). As most primate vigilance research has focused on the pre-emptive form of vigilance, we concentrate exclusively on it here.

In this study we identify four distinct definitional ethoses within primate vigilance literature. Firstly, a number of definitions are descriptive in nature, using visual terminology to describe the behavioural markers of interest, e.g., “cautiously observing” (de Ruiter, 1986), “gazing into the distance”, “gazing/gazes fixed on the surrounding environment” (Kutsukake, 2006, 2007), “visual inspection of surrounding area” (Fragaszy, 1990). Secondly, a number of studies have used and adapted the postural framework (e.g., head movement) found commonly in bird literature (see (Allan & Hill, 2018)). Examples include, “movement of the head and/or eyes” (Gaynor & Cords, 2012), “continuous head movement of at least 45 degrees in any direction” (MacIntosh & Sicotte, 2009), and “sweeping/single movement of the head” (Barros et al., 2008; Nunes et al., 2010). A third type of definition uses a non-operationalised scanning/looking approach, recording a basic posture and inferring nothing about the animal’s line of vision, e.g., “eyes were open and its head up” (Cowlishaw, 1998; Hill & Cowlishaw, 2002).

Finally, operationalised scanning/looking definitions have also been used. For example, Treves’ (e.g., (Treves, 1998)) scanning definitions “scanning/visual search directed beyond arm’s reach” or Allan & Hill's (2018) looking definition “Its eyes are open, and its line of vision extends beyond its hands and the substrate, animal or object that they are in contact with” ask observers to interpret when the animal is looking beyond its immediate vicinity, assuming this increases the animal’s chances of detecting a predator regardless of what it’s actually looking for or at. This definitional ethos is not focused on sampling vigilance specifically (see Treves, 1998 and Allan & Hill, 2018 for discussion), instead, the aim is to sample whenever an animal’s field of vision is such that it could detect a predator if it was there, regardless of its precise focus of attention. As such, much more general looking/scanning behaviours are recorded under the premise that scanning/looking and threat detection share complete compatibility. Studies adopting the definition of Treves (1998) have still frequently reported evidence regarding numerous vigilance hypotheses, highlighting that vigilance can still be detected analytically despite making no attempt to specifically sample a state of vigilance (see Allan & Hill, 2018).

We used a group of habituated Afromontane chacma baboons (*Papio ursinus griseipes*) as a model system to explore how the variation in vigilance definitions found in primate vigilance studies may impact on data consistency and repeatability. We constructed four representative definitions based on the definitional ethoses above to test the hypothesis that different definitions may contain inherent variability relating to their inter-observer reliability. This also allowed us to assess whether differences in vigilance definition can then lead to varied results. The definitions were:

1) **Visual terminology**: “Animal is gazing or visually inspecting its surroundings”
2) **Head/eye movement**: “Animal’s head is up combined with side-to-side movement of the head and/or eyes”
3) **Non-operationalised looking/scanning:** “It’s head is up and eyes open”
4) **Operationalised looking/scanning**: “Its eyes are open, and its line of vision extends beyond its hands and the substrate, animal or object that they are in contact with”

Predictions are challenging. Concise definitions (e.g., head up, eyes open) or definitions using verbs to describe behavioral/biological markers (e.g., gazing or inspection) could generate a greater degree of inter-observer variance. Alternatively, more complex operationalised definitions could result in observers struggling to converge on identical interpretations. Regardless, we investigate the reliability of the definitions themselves via inter-observer/rater agreement tests. An important issue when using experienced researchers to collect behavioural data is that they may have previously used a specific protocol, or several protocols, in the past. As such, each individual observer may be influenced by their initial and ongoing training interacting with their own experiences. We explore this notion in this study by using two separate groups of ‘experienced’ and ‘inexperienced’ observers.

Each definition and the unique interpretation of each definition by each observer may also produce differences in the distribution of datasets generated, and we assessed the magnitude of this variation and investigated whether it could influence the outcome of the results following a typical mixed-model vigilance analysis. We explored whether a number of widely investigated contextual factors (see Allan & Hill 2018) could influence looking/vigilance patterns differently depending on the definition used. We included number of neighbours within 5-meters, distance to nearest neighbour, habitat type (open/closed) and spatial position (central or peripheral) as these can also tie into hypotheses related to within-group and external threats (Allan & Hill, 2018). To incorporate the foraging-vigilance trade-off, we also included the time the animal spent performing 'engaged' behaviours, i.e., foraging and grooming.

This analysis is not focused on observer effects per se (i.e., the differences between observers within a study, or the impact of an observer on an animal's behaviour) but instead focuses on definition effects (i.e., differences in results due to the use of different definitions) and interpretation effects (i.e., differences in results between studies due to different interpretations of the same definition). Definition effects are unlikely to occur within a single study as researchers should not adopt multiple definitions of the same behaviour. Instead, definition effects are most likely to occur between different studies. If definition effects are apparent in this study, then we would expect varying results within observers and across definitions. Interpretation effects can overlap with observer effects within studies (e.g., several observers collecting vigilance data at the same time), but the focus of this study is to mimic occasions where multiple independent studies adopt similar definitions, and to explore whether the results are comparable in these instances. If interpretation effects are apparent in this study, then we would expect varying results within definitions and across observers.

Methods

Study area

This research was undertaken under ZA/LP/81996 research permit, with ethical approval from the Animal Welfare Ethical Review Board (AWERB) at Durham University. We collected our data on a wild habituated group of chacma baboons at Lajuma Research Centre in the western Soutpansberg

Mountains of Limpopo, South Africa (23°02’S 29°26’E). The study area included a complex mosaic of habitats within a variable mountainous environment (Willems & Hill, 2009) and all the natural habitats belonged to the Afromontane mist-belt communities and varied in their structural characteristics, including canopy height, foliage density and refuge availability (Coleman & Hill, 2014). The majority of the land within the study area was classified as a private nature reserve; however, agricultural practices take place locally and overlap with the core part of the study group’s home range (Williams et al., 2017). Known predators of the baboons in the study area include leopards and rock python, whilst the study group have also been observed to act fearfully and alarm at a number of raptor species and brown hyena.

Study group

Due to long-term anthropogenic activities in the study area (local farming and residences), consistent interactions with humans have been ongoing with this population for some time. The study group was formerly habituated circa 2005 and received periods of research attention up until 2014. Since 2014 the study group received consistent observational research in the form of full day follows (generally 3 to 4 days a week), with occasional gaps of up to 5 weeks duration. The group was typically followed dawn-to-dusk on a 4 days on 3 days off schedule designed to maintain as much of their natural interactions with predators as possible. The study group contained between 85 and 90 individuals over the course of the study (June 2018 to December 2018), several disappearances occurred during this period, causes were unconfirmed. AA had followed and collected behavioural and spatial data on this group since early 2015 and was able to identify all individuals including juveniles and infants based on their unique physical characteristics.

Video sampling methodology

All focal samples were completed solely by AA using a high-definition video camera (Panasonic HC-W580 Camcorder) to record all focal observations. Continuous focal sampling is the only method to capture detailed information on the temporal organisation of vigilance/looking (Allan & Hill, 2018;

Beauchamp, 2015; McVean & Haddlesey, 1980) and so we focused only on comparisons using a continuous focal sampling framework. Following pilot-study data exploring the ideal focal observation length for this study group, we used 30-second continuous focal animal sampling to collect vigilance/looking data across the full range of behaviours and habitat types. These short duration focal samples were appropriate for sampling the study animals' visual behaviours since the average bout length was often less than 1 second and the duration of the focal observation was rarely the same duration as time spent looking/vigilant. Across experienced observers, definition (1) produced 11.8% of samples with the duration of 'vigilance' equal to the observation length, whilst 2.2% of samples contained 0 seconds of vigilance, and 14% of samples had an average bout length of less than 1 second. The respective information for the remaining definitions was as follows: definition (2): 8%, 17.8%, and 4.2%; definition (3): 11.8%, 2.6%, and 15.2%; and definition (4): 12.4%, 0%, and 13.8%.

Short sampling periods are also an effective method to minimise the likelihood of aborted samples, require contextual variables to be updated less frequently, and reduce observer fatigue. When contextual factors are updated frequently it becomes challenging to manipulate and code data in a way that effectively explains the scenarios underpinning the focal observation, as such, short focal observations are an ideal solution to identify the precise drivers influencing the focal animal's current behaviour. Due to the difficulties associated with continuous focal sampling of moving animals a number of studies have excluded travelling activities (Gaynor & Cords, 2012; Smith et al., 2004; Stojan-Dolar & Heymann, 2010; Treves, 1998, 1999; Treves et al., 2001, 2003). As there is no prior expectation that baboons cannot detect threats if moving or do not collect visual information during travelling behaviours, it could form an important component of their looking repertoire, and thus was retained.

To control for time of day, we split the day into four time-periods that were adjusted seasonally to ensure each account for 25% of the current day length. We produced a randomly ordered list of all individuals in the group (excluding neonates and dependent infants) and selected focal individuals pseudo-randomly. The first individual encountered from the top 15 names (approx. 20% of original group-size) on the list was sampled immediately by AA. If greater than 50% of the focal animal's face was out of sight for more than 5 seconds, the focal observation was aborted. AA would then adjust his position and attempt to re-start the focal observation after a 1-minute break. This process was attempted a maximum of three times, after which AA would move on to sampling another individual from the list. The aborted focal individual was then reintegrated at the end of the list.

Prior training of 'experienced' and 'inexperienced' observers

Using the focal video observations collected solely by AA, we asked an additional nine observers (plus AA) to code these observations according to each definition. Four experienced observers (excluding AA) were all previously trained to collect 'scanning' data on habituated samango monkeys (*Cercopithicus albogularis schwarzi*) in field conditions using the definition "scanning directed beyond arm's reach" (see Treves, 1998). Each of these observers received identical training and testing and had completed at least 2 months of behavioural data collection in the field prior to the start of this study. This was the first primate behaviour and vigilance sampling experience each observer had received, so background knowledge should have been similar. The experience of sampling the definition of Treves (1998) may have interacted with the interpretation (of each definition) by this pool of observers and is very similar to the operationalised looking definition (definition iv) used in this study. To explore this, we enlisted a further 5 'inexperienced' observers (no previous experience studying vigilance in any organism, no previous experience collecting behavioural data on primates).

Extracting vigilance information from video-footage

We restricted the analysis to data collected on 18 individual adult females to limit the number of variables influencing the results; sex differences in vigilance and ontological effects have been reported in primates (see (Allan & Hill, 2018)). Looking/vigilance data was extracted from focal videos independently by the ten different observers (including AA) using the video playback software Media Player Classic (MPC-HC: Guliverkli project). Videos could be slowed down to extract precise looking bout lengths (video skip length could be reduced to 4 hundredths of a second when played back at quarter-speed). The start and end time for each looking/vigilance bout was ascertained from the media player and entered into an Excel spreadsheet, allowing two dependent variables to be calculated post-hoc for analysis: duration of time spent looking/vigilant and frequency of looking/vigilance bouts. Each focal video observation could be viewed as many times as needed, allowing precise information to be recorded. If the focal animal was deemed to be vigilant/looking at the onset of the focal observation, then bouts were considered to start with the commencement of the observation period. Similarly, an ongoing bout would be deemed to end at the end of the focal observation. If an observer felt that less than 50% of the focal animal's face was out of sight, this was coded as time spent out of sight and offset in model analyses (see below).

Video-coding regiment for observers

Each observer coded a number of focal sample videos using the different vigilance definitions. In order to make comparisons as robust as possible, all data was extracted from the same set of videos. The experienced group (plus AA) coded 10 focal videos for 18 individual female baboons (180 focal videos in total) for each definition. For the inexperienced group, we used 8 videos from a smaller subset of 8 individuals taken from the same dataset (64 videos coded in total) for each definition; these had also been coded by the experienced observers, allowing for comparisons. To act as a baseline for comparison, the inexperienced group were first asked to sample when they felt the animal was 'vigilant'. No further description was provided, or discussion of what vigilance may or may not mean. As such, we could assess the agreement for 'vigilance' within this group independent of definitions given subsequently.

The study was split into four distinct phases. Within each phase an observer was asked to extract data for a specific definition. To minimize a number of biases, each observer ran through the entire set of videos once for a single definition, before beginning to extract data for a different definition. Each observer was given a single definition at a time and asked to make their own interpretation of the definition before beginning the coding process and did not receive the next definition until coding the previous definition was complete. Observers were each given the definitions in a different order to mitigate against order effects, although the focal videos themselves were generally done in the same order (although an observer could choose to go through videos multiple times, so there may have been some sequencing discrepancy within definitions as a result). They could not return to another definition once it was completed.

Study authors did not guide observers towards specific interpretations. Our reason for doing this was to mimic how researchers may adopt the vigilance definitions of other researchers when replicating a study, thus, providing insight into whether interpretation effects could exist between independent studies. Some guidance was necessary however to ensure each observer created precise interpretations (i.e., based on their favoured behavioural, postural, or visual markers) that were consistently implemented across all observations. AA therefore encouraged all observers to think about the true meaning of each term or phrase (e.g., gazing, inspection, side to side, line of vision etc.), whilst additionally offering a range of postural (e.g., angle or height of the head from the ground), behavioural (e.g., raising or turning of the head), and visual cues (e.g., eye movement, direction of vision) for them to consider in their interpretations. All suggested cues were clearly observable as opposed to factors linked with the internal state of an animal, such as scanning intently, cautiously observing, or watchfulness, which have been used in primate vigilance research (see Allan & Hill 2018. AA used the same standard advice for all observers and kept a record of the cues each observer used, and the challenges they communicated whilst implementing each definition.

Each observer extracted data four times from each focal video, once for each definition given previously. The inexperienced group also extracted data for the fifth definition, 'vigilance', prior to the other four definitions. All observers agreed not to discuss their observations during each coding phase, or the duration of the study and no raters were able to code videos concurrently when in the same room. AA coded videos separate to the other raters and was the only participant not blind to the study design.

Contextual variables

While videoing focal observations in the field, AA recorded a number of contextual factors at the beginning and end of the 30-second focal observation. We use number of neighbouring conspecifics within 5 meters, distance to nearest neighbour, habitat type (open/closed) and spatial position (central or peripheral) as contextual variables that may predict vigilance use in baboons (Allan & Hill, 2018). The number of (all non-infant) neighbours or distance to nearest neighbour values were averaged for each focal between the start value (at zero seconds) and end value (at 30 seconds/end of focal). AA had validated their ability to assess both distance measures during pilot work but a calibrated laser range finder (Leica DISTO DXT) was used to assess distance to nearest neighbour if there were ever any accuracy concerns.

Habitat type and spatial position were assessed at the end of the focal. Habitat type was considered open when the focal animal was in areas without canopy cover (e.g., grassland, rock, cliff, marshland, road, camps, farm), and considered closed when canopy was present (e.g., bushland, woodland, forest). Spatial position of the focal animal was determined via assessment of visual and audible cues given by other group members. An individual was considered peripheral if on the edge of the group or had no more than 5 non-infant individuals between itself and the edge of the group. We used the focal videos to record the duration of engaged (foraging, grooming, self-grooming, handling food items) and not engaged (resting, moving, receive grooming, chewing food items) behaviours during a focal observation; and included this as another covariate predictor of vigilance use.

Calculating inter-observer/rater reliability (IRR) using concordance correlation coefficient (CCC)

Assessing IRR has traditionally used Pearson's correlation coefficients, paired t-tests, or Bland-Altman plots, but the intraclass correlation coefficient (ICC) has become more popular recently, especially when assessing reliability on continuous variables using multiple observers (Hernaez, 2015; Koo & Li, 2016). As the majority of our datasets did not meet the assumptions of ICC analysis, we instead calculated concordance correlation coefficients (CCC), estimated using a variance components analysis (see (Carrasco et al., 2013)). Concordance correlation coefficient is recommended in this scenario as it does not require the ANOVA assumptions of ICC analysis (Chen & Barnhart, 2013) and was shown to be identical to ICC when observers were treated as a fixed effect and agreement between ratings was being investigated (Carrasco & Jover, 2003)**.**

The data extraction protocol allowed for duration and frequency of bouts to be calculated for each focal observation. The CCC analysis was undertaken for both measures separately using a range of different groupings of observers. We first grouped all experienced observers together, with and without data produced by one author (AA). AA had completed several years of observational data collection on 3 monkey species and sampled several different vigilance or scanning definitions through this time and therefore had a different background to other experienced observers and was also aware of the purpose of the investigation. We grouped the inexperienced observers together (separately to experienced observers) as they were the only group to assess 'vigilance' and their CCC estimates were not as influenced by their training background and prior experiences. Finally, we grouped all observer's data together, with and without AA. As inexperienced observers coded a subset of data from the larger dataset, these CCC estimates were based only on focal observations both sets of observers had coded, i.e., the 64 videos coded by the inexperienced observers. Grouping all observers together allowed insights into whether the experienced and inexperienced groups produced consistent data to one another, as opposed to exploring the consistency within each observer grouping.

We calculated CCC using the “cccvc” function from the “cccrm” package (version 1.2.1) (Carrasco & Martinez, 2015), using the software R (version 3.5.0) (R Core Team, 2019). Since the focus of the study was to understand reliability within definition types, we did not require observers to code dummy datasets prior to this study and achieve an *a-priori* level of IRR before starting this study (Hallgren, 2012); the training aspect was simply to inform each observer of the extraction methodology and detail how data was required to be entered. All observers made assessments for every focal video on every individual baboon, making this study a fully crossed design (Hallgren, 2012).

The variance components model used for CCC estimation calculates the mean deviation of each observer from the overall mean across subjects and observers (Carrasco et al., 2013); in ICC terms this equates to the mean being used as the assessment basis for CCC estimation (Koo & Li, 2016). In ICC analyses a definition must be selected depending on whether absolute ‘agreement’ or ‘consistency’ are to be investigated. We designed this study to explore agreement between observers only, i.e., we were interested in whether multiple observers can produce similar values, as opposed to ‘consistency’ which tests whether observers’ ratings tend to produce similar rank orders (Hallgren, 2012; Koo & Li, 2016).

Focal observations with fewer bouts or less time devoted to vigilance or looking activities clearly have less potential for discrepancy between observers than focal observations where vigilance/looking is frequent. In addition, when values for either dependent variable was lower, small discrepancies between observers could lead to higher variability between them relative to discrepancies between observations with higher observed levels of vigilance/looking. However, these factors are not issues within this analysis as all observers coded the same set of video observations (i.e., fully crossed design) across a diverse range of looking/vigilance values. As such, to get excellent agreement (i.e., CCC > 0.9) then almost identical assessments are required across observers (within observations) for a high proportion of observations.

Mixed model analysis assessing the consistency of trends across definitions and observers

The data produced for each definition by each ‘experienced’ observer plus AA was used for this aspect of the analysis. Duration and frequency measures (produced from each definition, by each observer) were used as separate response variables in a number of generalized linear mixed effects models; each observer had four models with duration as a response variable, and four models with frequency as the response variable, each based on the data for the four definitions coded. Each model was fitted using a Bayesian procedure and the same contextual variable predictors. The time the focal animal had at least 50% of its face in view (never less than 25 seconds) was included as an offset variable in all models, e.g., if the animal’s face was out of view for two seconds then the observation length was updated to 28 seconds within the offset variable. Individual baboon ID was fit as a random factor. In all cases the error distribution for the duration (seconds) response models was Gaussian and Poisson for the frequency (count) response models, each with default link functions. Default half student t priors (df = 3, M = 0, SD = 10) were assigned for fixed and random effects within the *brm* function.

All models were fit using the *brm* function from the brms package (Bürkner, 2017) in the R software (R Core Team, 2019). The *brm* function commands samples to be drawn from the posterior distribution via the package Rstan (Team, 2019a), which interfaces with the probabilistic programming language Stan (Team, 2019b) via the C++ toolchain in Rtools (R Core Team, 2018). The *brm* function implements Hamiltonian Monte Carlo (Duane et al., 1987; Neal, 2003) in combination with the No-U-Turn Sampler (NUTS) extension (Hoffman & Gelman, 2014), resulting in algorithms that converge efficiently, even for complex models (Hoffman & Gelman, 2014). For each model we ran six Hamiltonian Markov chains for 10000 iterations to provide algorithms which converge efficiently for multi-level models (Bürkner, 2017); these were both set higher than default settings to aid fitting a relatively small sample size. In addition, we set warmup to 4000 (higher than default). This sets the number of warmup iterations used for stepwise adaptation and allows the sampling algorithm to hone in on efficient values for step size and the number of steps used for sampling (Bürkner, 2017; McElreath, 2019). Warmup iterations were discarded and not used for sampling; using a higher warmup than default improves sampling efficiency and aids in modelling of the entire posterior distribution including potentially extreme tails (McElreath, 2019)**.**

To aid with issues relating to a small sample size Adapt_delta was set to 0.99; this reduces the step size (which controls the resolution of the NUTS sampler) forcing the NUTS sampler to slow down, producing more robust posterior samples. Across all models there was no evidence of divergent transitions. The Gelman-Rubin convergence diagnostic (Rhat) (Gelman & Rubin, 1992) was used to assess Markov Chain Monte Carlo (MCMC) convergences by comparing the estimated within- and between-chain variances of each factor within the model. Rhat was equal to 1 in all cases, indicating accuracy of the response variables with regards to the Gaussian/Poisson response distributions, i.e., the standard deviation of duration/frequency points formed around the corresponding Guassian/Poisson functions was minimal. In all models, the bulk and tail estimated sample size (ESS) was greater than 10,000 for all fixed effects.

Although we principally examined the estimates, 95% credible intervals, and conditional effects from each model, this process can be subjective and lead to incorrect interpretations of results (Kruschke, 2018). Therefore, we additionally calculated the 89% Highest Density Interval (HDI) of the posterior distribution of each model. The HDI reveals the upper and lower parameter values of the posterior distribution based on all points within the 89% interval, points within the interval therefore have a higher probability density than points outside the interval (Kruschke & Liddell, 2018). Although any arbitrary percentage value could be implemented for the HDI, 89% has been recommended due to it providing improved stability over using 95% (McElreath, 2019). We also calculated the probability of direction (pd) for each fixed effect for each model. The pd variable is an index for inspecting effect existence and highlights the certainty that a particular effect has directionality (i.e., is positive or negative); pd ranges from 50% (i.e., equal distribution of positive and negative posterior values) to

100% (e.g., all posterior values are either positive or negative), pd has additionally been shown to have a 1:1 correspondence with p-values derived using frequentist methods (Makowski et al., 2019).

In order to reduce potential for inconsistent interpretations of results we developed an *a-priori* set of rules for deducing results. We used two criteria for accepting the null hypothesis, 1) when the HDI overlapped or included zero, and 2) when the pd was less than 90%. This would mean at least a proportion of the most credible parameter values include zero whilst pd indicates little certainty in the effect having directionality. In all other circumstances the null was rejected, and we classified results as 'effect has some uncertainty', 'moderate evidence for an effect', and 'strong evidence for an effect'. In all of these scenarios there needed to be some evidence for a relationship between a covariate and the dependent variable, i.e., a positive/negative estimate, evidence of a consistent relationship in conditional effects plots. Moderate evidence for an effect required that the HDI did not overlap or include zero and that the pd was at least 90%, the only difference to strong evidence was that the pd was higher than 97.5%. There were numerous cases of detecting positive/negative estimates with noticeable trends between the covariate predictors and the dependent variable, and the pd was at least 90 or 95%, but the HDI marginally overlapped or included zero. These cases still suggest some evidence of an effect, but there was uncertainty, and these results were coded accordingly. All estimates and HDI values displayed in the results tables were coded according to the HDI-pd decision rules (see table 2 and 3) to aid visual interpretation.

Results

Inter-observer reliability (IRR)

The experienced group produced excellent agreement for duration of vigilance/looking for 2 out of 4 definitions (see table 1, definitions: 1 and 4), with definition (3) approaching excellent agreement. The effect was reduced for frequency estimates, with only definition (4) approaching excellent agreement. The operationalised definition (4) seems to be the most repeatable for both duration and frequency, although the visual terminology definition (1) is close amongst experienced observers, particularly for duration assessments. The inexperienced group produced almost identical agreement results (compared to experienced) for duration and frequency for definition (4), this held when their data was combined with experienced observers, including and not including AA. For inexperienced observers, agreement was moderate for duration of definition (1), estimates were considered poor for (2) and (3), although definition (3) was moderate for combined assessments. Agreement was moderate (for duration) amongst inexperienced observers for the 'vigilance' definition, which was greater than the agreement they produced for definitions (2) and (3). Agreement was poor (for frequency) amongst inexperienced observers for 4 out of 5 definitions, but was again slightly higher for 'vigilance' than for definitions (1, 2 , 3).

[Table 1 here]

Consistency of trends across definitions and observers

The mixed model analysis assessing the consistency of trends with contextual variables for the duration measure found examples of results remaining consistent in direction and magnitude within observers across definitions (see table 2). However, there were also a number of examples of results differing. For example, observer 1's data found evidence of an effect for time spent engaged in definitions (1) and (4), but this effect had uncertainty in definition (3), whilst the H0 was accepted in definition (2). Generally, each definition produced similar directionality of results for the duration response variable, although a small number of cases reported an opposing trend. Model estimates and HDI parameter values were generally consistent across observers for definitions (1) and (3), although both have examples of varied results. Definition (4) was also consistent and found similar estimates and HDI parameter values across all variables; however, the uncertainty around some results could lead to varied interpretations. Definition (2) produced the most varied results with the spatial position variable producing both strong and moderate evidence for three observers but accepting the H0 in the remaining two.

[Table 2 here]

Agreement of results was lower for the frequency measure compared to the duration measure (table 3), with variable results within observers across definitions and within definitions across observers. The model estimates and HDI parameter values were fairly consistent for the engaged variable, however, some minor differences were apparent. The number of neighbours within 5 meters variable produced consistent directionality, however, there was some variation in estimates and HDI parameter values across definitions and observers, despite this definition (1) and (4) produced the most consistent results. It could be argued that the average nearest neighbour distance generally produced posterior values relatively close to zero and as such there is little evidence supporting an effect in any model. However, several models display a pd of greater than 97.5%, indicating strong evidence supporting a negative effect of average nearest neighbour distance; as such, results clearly varied across definitions and observers. The habitat (open/closed) variable consistently produced posterior estimates that supported the null for definitions (1), (2), and (3); however, for definition (4) three observer's models found moderate or strong evidence supporting a negative effect. Spatial position (central/peripheral) was also relatively consistent in directionality with the H0 accepted for most models, although there was some evidence supporting an effect in the results produced from AA's models for definition (1) and (2).

For frequency assessments there appears to be strong within definition agreement between some observers but not others. Examples include, AA, observer 1, and observer 3 producing almost identical posterior estimates and HDI parameter values for definition (4) whilst observers 2 and 4 produced less substantial evidence. Observers 2 and 3 produced similar estimates and HDI parameter values for definitions (1) and (2) but varied for definitions (3) and (4), whilst AA and observer 1 were consistent for definitions (3) and (4), and observers 2 and 4 produced similar findings across all models. In summary for frequency assessments, each definition produced generally varied results within observers, whilst there wasn't a single definition that produced identical results across observers, even if decision rules were to be relaxed**.**

[Table 3 here]

Discussion

We found variation in inter-observer agreement across four different types of vigilance definition used in primate research. When viewing results across all definitions and experienced/ inexperienced observers we found excellent agreement within definition (4) operationalised looking, for both duration and frequency assessments, suggesting this definition is capturing similar information across observers. Agreement was found for other definitions in certain scenarios, but this did not carry through into frequency assessments or across the experienced/inexperienced dichotomy. Comparative model results for data produced by experienced observers suggested that each definition could lead to different results, which could vary across observers, supporting notions of definition and interpretation effects. However, definition (4) produced slightly more consistent results, with similar posterior values produced across observers for duration assessments. In general, model results for frequency assessments were more variable than for duration. Definition (4) produced very consistent estimates and HDI parameters values for 3 observers, but did not produce identical results across all observers, suggesting interpretation effects may be more important if frequency assessments are being investigated. Together these results suggest that while operationalised definitions produce the most consistent results for baboons, authors must take care when selecting or constructing new definitions for future work and that issues may vary according to the choice of dependent variable.

Our study setup was designed to minimise the amount of variation that was likely to occur due to observer experience and numerous methodological factors. Observers were afforded the time to independently specify their observations with accuracy. Making similar assessments in the field using behavioural software on mobile devices or stop watches/clickers (amongst other methods) is likely to be less precise (and risks missing bouts entirely), with no possibility for observers to rectify accidental mistakes. The use of video-coding techniques likely plays an important role in observer interpretation of definitions and is itself vital methodological information. For example, each observer assessed numerous bouts that lasted less than half a second, this would be impossible to implement precisely in real-time/field conditions and thus would impact on how an observer interpreted and implemented a definition. Consequently, our results are at the lower end of the variation we might expect in natural settings where observers only view the behaviour once. Despite controlling for these factors, we still found variation in consistency within definitions and variation in results across definitions, suggesting there are important implications of definitions that researchers must be aware of.

Our methods controlled for other sources of variation that are likely to be important in the literature, such as variation in study species, sampling methodology (Hirschler et al., 2016), dependent variables (e.g., average bout length, inter-scan interval, proportion of time vigilant/looking), and statistical procedures (see supplementary material Allan & Hill 2018). Interactions between these factors and sampling methodology are critical as some definitions should not be possible with instantaneous point sampling; for example, definition (2) requires movement through time to be assessed. Definitions such as definition (4) may require extensive training periods to refine an observer's search images before collecting data, and even then, assessments may be more challenging with instantaneous point sampling versus continuous or one-zero sampling. Observer fatigue is another factor we had some control over in this study as our observations were a maximum of 30 seconds duration and observers could take breaks as frequently as needed from video-coding. In field conditions, interpretation and definition effects may be amplified according to various challenges including focal durations, observation day length, climatic/weather conditions, and any factors relating to the observer's own aptitude, attitude, and emotional state.

We suggest researchers adopt video-sampling methods, whether recorded directly by an observer or via remote technologies such as camera traps. This should improve the precision and accuracy of observations, whilst offering the advantage of preserving observations, affording researchers the opportunity to apply alternative sampling methods post-hoc. Videos may also be useful as new computational methods (e.g., computer tracking of head angles or line of vision) are developed, which may offer solutions to interpretation effects. In addition, videos offer authors the opportunity to monitor other observer's assessments through time to guard against interpretation and precision issues.

A limitation here may be that using 'experienced' observers that were familiar with different sampling methods and vigilance definitions previously could also have an important interaction with interpretation effects. Amongst experienced observers, agreement was excellent for definitions (1) and (4) (for the duration variable) and was approaching excellent for definition (3). Agreement was substantially lower for definition (1) and (3) across inexperienced observers and when all observers' data was grouped. Prior experience of training, testing, and observations using the definition of Treves (1998) ('scanning/visual search directed beyond arm's reach'), may have biased experienced observers to interpret definitions (1), (3), and (4) similarly to Treves' definition, and by extension similar to one another. This seems evident as the agreement results of inexperienced observers highlighted that definition (4) was the only definition to maintain excellent agreement across all groupings, whilst definitions (1) and (3) achieved moderate and poor agreement respectively. It would be interesting to explore these factors in future research as researchers who have worked on studies using certain definitions could potentially produce interpretational effects when implementing new definitions that are different to their previous work. These findings would also generalise to behaviours other than vigilance where definitions differ between studies.

Frequency assessments generally provided lower agreement, but definition (4) produced agreement estimates close to excellent regardless of how the data was grouped. These results suggest definition (4) may be somewhat more robust to prior training and experiences and could aid in making cross-study comparisons reliable. However, this may have been driven by exceptional agreement between AA, observer 1, and observer 3, as the model estimates and HDI parameter values of their models were almost identical, whilst observers 2 and 4 did not share similar findings. This suggests that even when high inter-observer reliability is found, minor differences in interpretation can still produce different results, which could have important implications relating to the outcome of inter-rater reliability/consistency tests in observational studies.

It seems likely that the definition presented by Treves (1998) “scanning/visual search directed beyond arm’s reach”, would perform similarly to looking (4) “eyes are open, and its line of vision extends beyond its hands and the substrate, animal or object that they are in contact with”, in terms of inter-observer agreement and consistency in results. Even so, this should be formally tested before assumptions are made, as our results suggest that even minor sources of variation can influence the direction and magnitude of results. One key difference between the two definitions is that ‘scanning (or visual search) directed beyond arm’s reach’ could suggest a deliberate form of visual information acquisition, which may be interpreted as requiring some form of ‘active’ scanning. As such, all observers may not sample animals simply resting with their eyes open consistently. Over several years of training this definition with numerous observers (and primate species), AA found this to be a consistent source of inter-observer discrepancy in interpretation; ultimately, motivating the decision to present the operationalised looking definition (see Allan & Hill 2018). The looking definition instead focuses on the animal’s ‘line of vision’ as opposed to ‘scanning’ or ‘visual search’, which should allow for the unanimous inclusion of passive bouts of looking and produce more consistent data across observers than the scanning definition.

The behavioural variable (engaged/not engaged) was the only variable that could theoretically have interacted with the interpretations of each definition since certain definitions may be difficult to operationalise during engaged behaviours (e.g., foraging or grooming). However, behaviour produced fairly consistent model estimates and HDI parameter values across both duration and frequency measures; suggesting that behaviour was not a primary explanation of variation in interpreting definitions across observers. As contextual variables relate to the focal animal’s surrounding environment, they should not interact with the interpretation of each definition. This was supported for duration assessments regarding the spatial position variable, as model estimates and HDI parameter values were similar across definitions and observers although, the magnitude of posterior values was noticeably lower (i.e., closer to zero) for the models implemented using observer 1 and 2's duration assessments for definition (2), whilst observer 4 produced noticeably higher posterior values for definition (2) relative to other observers.

Results for the habitat variable were consistent for models using duration assessments, with all models accepting the null hypothesis. The null hypothesis was also accepted in all models for frequency assessments of definitions (1), (2), and (3); however, definition (4) produced evidence for an effect in three observers' models. Results were similarly consistent for the number of neighbours and distance to nearest neighbour variables for duration assessments, with only models produced by observer 2 yielding enough evidence to potentially support an effect for definitions (1) and (2). Results were more variable for number of neighbours and distance to nearest neighbour for frequency measures. Importantly, the two social environment variables produced results that varied with respect to one another, both within observers across definitions and within definitions across observers. This is significant since each of these factors generally represents the hypotheses under investigation in vigilance studies. The latter finding also suggests that given the array of methods for sampling contextual variables in primate vigilance research (see supplementary material Allan & Hill 2018), consolidating towards a common method of sampling certain contextual factors may be important going forward.

We believe our results highlight a set of phenomena that are often overlooked in observational research, namely definition effects and interpretation effects. Definitions effects are highlighted well with each observer producing varied results across definitions. Interpretation effects are also well supported in both our analyses. Firstly, definitions clearly have the potential to differ in inter-observer reliability and secondly, it is also clear that regardless of high inter-observer reliability, model estimates, HDI parameter values, and ultimately the results we interpret can vary substantially within definitions but across observers. Interpretation effects are most likely to manifest between independent studies using similar definitions. In these cases, differences in results could exist purely due to differences in the interpretations made by principal investigators; however, this could be further exacerbated if differences in interpretation between observers within the same study are also allowed to manifest. Studies using multiple observers during the same period can clearly control for some issues with appropriate training and testing programs, however, we show here that even excellent agreement results do not guard against interpretation effects between observers within the same study. Another option may be to include observer identity in statistical analyses; however, this has rarely been used in primate vigilance research (see supplementary material Allan & Hill, 2018)) and it is unclear whether this would adequately control for interpretation effects. We kept the study species the same throughout our study, but it seems likely definition and interpretation effects would be even more problematic when applied across different species and taxa.

It seems likely that certain ethoses are more prone to within-observer variation too, i.e., observers applying definitions inconsistently. We tried to avoid this source of variation in our study by actively encouraging assistants to take as much time as needed to refine their interpretations and code observations, however, most still found our set list of terms and phrases e.g., head/eye angle from the ground, head/eye movement, degree of head movement, useful when formulating their initial interpretations. It is possible that this process may have introduced some bias, however, it highlighted to us that all definitions bar the operationalised looking definition were initially challenging for our observers to confidently interpret and implement without offering some advice. We suggest that future research considers avoiding defining behaviours using terms such as gazing, scanning, inspection, or watchfulness, as we found these to be ambiguous and placed an onus on interpreting the internal state of an animal as opposed to assessing external markers. Our observers also found terms such as head up or side to side movement challenging to implement without further detail operationalizing when bouts begin and end, even then assessing head angles from the ground may be challenging to reach agreement on. The looking definition circumnavigated some of these interpretation issues by asking observers to focus on the animal's line of vision in relation to their hands, which is unlikely to yield a diverse range of interpretations given its operational nature.

The operational looking definition also offers an additional advantage as it likely maximizes the amount of information collected by researchers. The behavioural markers typically used to define vigilance in non-primate species have focused on postural changes (e.g., head raising) but some species may value visual information gained during 'head-down' postures (Bednekoff & Blumstein, 2009; Bednekoff & Lima, 2005), while the detective capabilities of other species may not be hindered during some foraging tasks (Allan et al., 2020; Kaby & Lind, 2003). These findings highlight the issue with using postural changes as markers for vigilance and measures for fearfulness (Tatte et al., 2019) as animals can achieve vigilance goals during several postures and engaged behaviours. We believe focusing on looking (i.e., when an animal's line of vision is unobstructed and angled away from their local vicinity) is a viable solution in species where line of vision can be assessed reliably. When this is not possible authors should consider consolidating towards a key set of unambiguous markers that allow a full range of visual behaviours to be sampled. In birds for example, the looking definition could be operationalised as any time the animal's field of view/vision is not obstructed within certain distances (e.g., a wing's length, body length, or a meter); such circumstances likely allow birds to collect information on their surrounding environment even during head-down postures or during foraging. Collection of this type of information would require additional work concerning the sensory capacity of study species during a full range of postures, behaviours, and scenarios (Allan & Hill, 2018). Nevertheless, however, it could yield a more complete understanding of how animals monitor their environment for risks. Our results highlight that sampling looking can still allow vigilance hypotheses to be tested, and risk drivers to be elucidated despite not sampling vigilance specifically.

To conclude, our findings suggest that behavioural data collection methods need to be as consistent as possible to allow for robust comparisons across study sites, species, and individuals. Although we find support for the looking definition in this study, our results also indicate that very minor differences in observer interpretation can lead to varied results. Nevertheless, without convergence towards a single definition it may still be challenging to compare results both within species across populations and studies and in comparative studies across species. In baboons, we believe operational definitions are a necessity and that looking behaviours are likely to capture the most amount of information towards understanding the functions of vigilance. Operationalised looking has the potential to apply to any species with forward-facing eyes but may apply elsewhere too, particularly other catarrhine species. We encourage researchers to explore designs similar to our own in other taxa and work together to develop a more complete understanding of the extent and solutions to these issues. Even without a universal definition, it would be useful to explore and debate consolidation onto similar definitions and methods within taxonomic families and to refine the daunting list of terms and phrases currently used to define vigilance behaviours (see Allan & Hill 2018). Ultimately, we hope our results are useful in instigating a wider debate amongst behavioural ecologists about definitions and sampling design for all behaviours, not just vigilance.

Acknowledgements

We send out a massive thank you to Aaron Eastwood, Amy White, Joe Jacobs, Rebekah Kelly, Ana Kovacic, Annie Powell, Bill Wang, Cyrintha Barwise-Joubert, and Eiylish Powell for extracting the data from the videos, following all the rules, and persevering through the study. We also thank Prof. Ian Gaigher and Jabu Linden for permission to conduct research on the Lajuma property, and the neighbouring landowners for access to their properties for data collection. We also thank Laura LaBarge for useful discussions throughout all stages of the project and for constructive comments on earlier versions of the manuscript. Finally, we also thank Professor Susan Healy and two anonymous reviewers for their feedback on our manuscript which we used to improve the final version.

Funding

AA was funded by a Natural Environment Research Council (NERC) studentship through the IAPETUS Doctoral Training Partnership.

References

Alberts, S. C. (1994). Vigilance in young baboons: effects of habitat, age, sex and maternal rank on glance rate. *Animal Behaviour*, *47*(4), 749–755.

Allan, A. T. L., Bailey, A. L., & Hill, R. A. (2020). Habituation is not neutral or equal: Individual differences in tolerance suggest an overlooked personality trait. *Science Advances*, *6*, eaaz0870.

Allan, A. T. L., & Hill, R. A. (2018). What have we been looking at? A call for consistency in studies of primate vigilance. *American Journal of Physical Anthropology*, *165*, 4–22.

Arenz, C., & Leger, D. (2000). Antipredator vigilance of juvenile and adult thirteen-lined ground squirrels and the role of nutritional need. *Animal Behaviour*, *59*(3), 535–541.

Atkins, A., Little, R. M., Redpath, S. M., & Amar, A. (2019). Impact of increased predation risk on vigilance behaviour in a gregarious waterfowl, the Egyptian goose *Alopochen aegyptiaca*. *Journal of Avian Biology*, *50*(e02121), 1–8.

Barros, M., Alencar, C., Silva, M. A. de S., & Tomaz, C. (2008). Changes in experimental conditions alter anti-predator vigilance and sequence predictability in captive marmosets. *Behavioural Processes*, *77*(3), 351–356.

Beauchamp, G. (2015). *Animal vigilance: monitoring predators and competitors*. Academic Press.

Beauchamp, G. (2017). What can vigilance tell us about fear? *Animal Sentience*, *2017.015*(Beauchamp on Fear & Vigilance), 1–53.

Beauchamp, G. (2018). The effect of age on vigilance: a longitudinal study with a precocial species. *Behaviour*, *155*(13), 1011–1024.

Bednekoff, P. A., & Blumstein, D. T. (2009). Peripheral obstructions influence marmot vigilance : integrating observational and experimental results. *Behavioral Ecology*, *20*, 1111–1117.

Bednekoff, P. A., & Lima, S. L. (2005). Testing for peripheral vigilance: Do birds value what they see when not overtly vigilant? *Animal Behaviour*, *69*(5), 1165–1171.

Blumstein, D. T., Lea, A. J., Olson, L. E., & Martin, J. G. A. (2010). Heritability of anti-predatory traits : vigilance and locomotor performance in marmots. *Journal of Evolutionary Biology*, *23*, 879–887.

Boinski, S. U. E., Kauffman, L., Westoll, A., Stickler, C. M., Cropp, S., & Ehmke, E. (2003). Are vigilance, risk from avian predators and group size consequences of habitat structure? A comparison of three species of squirrel monkey (*Saimiri oerstedii, S. boliviensis, S. sciureus*). *Behaviour*, *139*, 1421–1467.

Brandl, S. J., & Bellwood, D. R. (2015). Coordinated vigilance provides evidence for direct reciprocity in coral reef fishes. *Scientific Reports*, *5*(14556), 1–13.

Brown, N. A., Ruckstuhl, K. E., Donelon, S., & Corbett, C. (2010). Changes in vigilance, grazing behaviour and spatial distribution of bighorn sheep due to cattle presence in Sheep River Provincial Park, Alberta. *Agriculture, Ecosystems and Environment*, *135*, 226–231.

Burger, J., & Gochfeld, M. (1988). Effects of group size and sex on vigilance in ostriches (*Stuthio camelus*): Antipredator strategy or mate competition? *Ostrich: Journal of African Ornithology*, *59*(1), 14–20.

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Campos, F. A., & Fedigan, L. M. (2014). Spatial ecology of perceived predation risk and vigilance behavior in white-faced capuchins. *Behavioral Ecology*, *25*(3), 477–486.

Carrasco, J. L., & Jover, L. (2003). Estimating the Generalized Concordance Correlation Coefficient through Variance Components. *Biometrics*, *59*, 849–858.

Carrasco, J. L., & Martinez, J. P. (2015). *Package ' cccrm .'*

Carrasco, J. L., Phillips, B. R., Puig-Martinez, J., King, T. S., & Chinchilli, V. M. (2013). Estimation of the concordance correlation coefficient for repeated measures using SAS and R. *Computer Methods and Programs in Biomedicine*, *109*, 293–304.

Carter, A. J., Pays, O., & Goldizen, A. W. (2009). Individual variation in the relationship between vigilance and group size in eastern grey kangaroos. *Behavioral Ecology and Sociobiology*, *64*(2), 237–245.

Chen, C., & Barnhart, H. X. (2013). Assessing agreement with intraclass correlation coefficient and concordance correlation coefficient for data with repeated measures. *Computational Statistics and Data Analysis*, *60*, 132–145.

Coleman, B. T., & Hill, R. A. (2014). Living in a landscape of fear: The impact of predation, resource availability and habitat structure on primate range use. *Animal Behaviour*, *88*, 165–173.

Cowlishaw, G. (1998). The role of vigilance in the survival and reproductive strategies of Desert Baboons. *Behaviour*, *135*(4), 431–452.

Cowlishaw, G., Lawes, M. J., Lightbody, M., Martin, A., Pettifor, R., & Rowcliffe, J. M. (2004). A simple rule for the costs of vigilance: empirical evidence from a social forager. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1534), 27–33.

de Ruiter, J. R. (1986). The influence of group size on predator scanning and foraging behaviour of wedgecapped capuchin monkeys (*Cebus olivaceus*). *Behaviour*, *98*(1), 240–258.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216–222.

Elgar, M. A., & Catterall, C. P. (1981). Flocking and predator surveillance in house sparrows: Test of an hypothesis. *Animal Behaviour*, *29*(3), 868–872.

Favreau, F.-R., Goldizen, A. W., & Pays, O. (2010). Interactions among social monitoring, anti-predator vigilance and group size in eastern grey kangaroos. *Proceedings of the Royal Society B*, *277*, 2089–2095.

Fernandez, G. J., Capurro, A. F., & Reboreda, J. C. (2003). Effect of Group Size on Individual and Collective Vigilance in Greater Rheas. *Ethology*, *109*, 413–426.

Ferrari, C., Bogliani, G., & von Hardenberg, A. (2009). Alpine marmots (*Marmota marmota*) adjust vigilance behaviour according to environmental characteristics of their surrounding. *Ethology Ecology & Evolution*, *21*, 355–364.

Forslund, P. (1993). Vigilance in relation to brood size and predator abundance in the barnacle goose, *Branta leucopsis*. *Animal Behaviour*, *45*(5), 965–973.

Fragaszy, D. (1990). Sex and age differences in the organisation of behaviour in wedge-capped capuchins, *Cebus olivaceus*. *Behavioral Ecology*, *1*(1), 81–94.

Gaynor, K. M., & Cords, M. (2012). Antipredator and social monitoring functions of vigilance behaviour in blue monkeys. *Animal Behaviour*, *84*(3), 531–537.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Hernaez, R. (2015). Reliability and agreement studies: a guide for clinical investigators. *Gut*, *64*(7), 1018–1027.

Hill, R. A., & Cowlishaw, G. (2002). Foraging female baboons exhibit similar patterns of antipredator vigilance across two populations. In L. E. Miller (Ed.), *Eat or Be Eaten: Predator Sensitive Foraging*

Among Primates (pp. 187–204). Cambridge University Press.

Hirschler, I. M., Gedert, J. L., Majors, J., Townsend, T., & Hoogland, J. L. (2016). What is the best way to estimate vigilance? A comparison of two methods for Gunnison's prairie dogs, *Cynomys gunnisoni*. *Animal Behaviour*, *121*, 117–122.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Ito, R., & Mori, A. (2010). Vigilance against predators induced by eavesdropping on heterospecific alarm calls in a non-vocal lizard *Oplurus cuvieri cuvieri* (*Reptilia: Iguania*). *Proceedings of the Royal Society B*, *277*, 1275–1280.

Javier, D.-R. F., & Perez-Mellado, V. (2000). The effects of population density on time budgets of the iberian wall lizard (*Podarcis hispanica*). *Israel Journal of Zoology*, *46*, 215–229.

Kaby, U., & Lind, J. (2003). What limits predator detection in blue tits (*Parus caeruleus*): Posture, task or orientation? *Behavioral Ecology and Sociobiology*, *54*, 534–538.

Kahlert, J. (2003). The constraint on habitat use in wing-moulting Greylag Geese *Anser anser* caused by anti-predator displacements. *Ibis*, *145*, E45–E52.

Klett-mingo, J. I., Pavon, I., & Gil, D. (2016). Great tits, Parus major, increase vigilance time and reduce feeding effort during peaks of aircraft noise. *Animal Behaviour*, *115*, 29–34.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, 1–11.

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*, 155–177.

Kutsukake, N. (2006). The context and quality of social relationships affect vigilance behaviour in wild chimpanzees. *Ethology*, *112*(6), 581–591.

Kutsukake, N. (2007). Conspecific influences on vigilance behavior in wild chimpanzees. *International Journal of Primatology*, *28*(4), 907–918.

Lanham, E. J., & Bull, C. M. (2004). Enhanced vigilance in groups in *Egernia stokesii*, a lizard with stable social aggregations. *Journal of Zoology*, *263*, 95–99.

Li, D., Liu, Y., Sun, X., Lloyd, H., Zhu, S., Zhang, S., & Wan, D. (2017). Habitat-dependent changes in vigilance behaviour of Red-crowned Crane influenced by wildlife tourism. *Scientific Reports*, *7*(16614), 1–11.

Lian, X., Zhang, T., Cao, Y., Su, J., & Thirgood, S. (2007). Group size effects on foraging and vigilance in migratory Tibetan antelope. *Behavioural Processes*, *76*, 192–197.

Lopez, P., & Martin, J. (2013). Effects of Microhabitat-Dependent Predation Risk on Vigilance during Intermittent Locomotion in Psammodromus algirus Lizards. *Ethology*, *119*, 316–324.

MacIntosh, A. J. J., & Sicotte, P. (2009). Vigilance in ursine black and white colobus monkeys (*Colobus vellerosus*): An examination of the effects of conspecific threat and predation. *American Journal of Primatology*, *71*(11), 919–927.

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Ludecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Retrieved from Https://Doi.Org/10.31234/Osf.Io/2zex*, 1–35.

McElreath, R. (2019). *Statistical Rethinking 2: A Bayesian Course with Examples in R and Stan* (Second).

McVean, A., & Haddlesey, P. (1980). Vigilance schedules among House Sparrows *Passer domesticus*. *Ibis*, *122*, 533–536.

Neal, R. M. (2003). Slice sampling: Rejoinder. *Annals of Statistics*, *31*(3), 705–767.

Nunes, D. M., Gonçalves, I., Emile, N., & Barros, M. (2010). Bimodal temporal organization of specific vigilance behaviors in captive black tufted-ear marmosets (*Callithrix penicillata*). *Behavioural Processes*, *84*(2), 629–631.

Pulliam, H. R. (1973). On the advantages of flocking. *Journal of Theoretical Biology*, *38*, 419–422.

Quenette, P. (1990). Functions of vigilance behavior in mammals: a review. *Acta Oecologica*, *11*(6), 801–818.

Quirici, V., Castro, R. A., Oyarzún, J., & Ebensperger, L. A. (2008). Female degus (*Octodon degus*) monitor their environment while foraging socially. *Animal Cognition*, *11*(3), 441–448.

R Core Team. (2018). *Rtools software, Version 35*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team. (2019). *R: A language and environment for statistical computing. Version 3.6.1*. R Foundation for Statistical Computing, Vienna, Austria.

Rieucau, G., & Martin, J. G. A. (2008). Many eyes or many ewes: vigilance tactics in female bighorn sheep *Ovis canadensis* vary according to reproductive status. *Oikos*, *117*, 501–506.

Shimada, T., & Shimada, K. (2003). Factors affecting vigilance in Greater White-fronted Geese *Anser albifrons*. *Wildfowl*, *54*, 201–208.

Shriner, W. M. (1998). Yellow-bellied marmot and golden-mantled ground squirrel responses to heterospecific alarm calls. *Animal Behaviour*, *55*, 529–536.

Smith, A. C., Kelez, S., & Buchanan-Smith, H. M. (2004). Factors affecting vigilance within wild mixed-species troops of saddleback (*Saguinus fuscicollis*) and moustached tamarins (*S. mystax*). *Behavioral Ecology and Sociobiology*, *56*(1), 18–25.

Stojan-Dolar, M., & Heymann, E. W. (2010). Vigilance in a cooperatively breeding primate. *International Journal of Primatology*, *31*(1), 95–116.

Tadeo, M. P., & Gammell, M. (2018). Activity budgets in different habitats of a species of conservation concern in Ireland, the Light-bellied Brent Goose *Branta bernicla hrota*. *Wildfowl*, *68*, 84–103.

Tatte, K., Ibanez-Alamo, J. D., Marko, G., Mand, R., & Moller, A. P. (2019). Antipredator function of vigilance re-examined: vigilant birds delay escape b. *Animal Behaviour*, *156*, 97–110.

Team, S. D. (2019a). *RStan: The R interface to Stan*. R Foundation for Statistical Computing, Vienna, Austria.

Team, S. D. (2019b). *The Stan C++ Libary, Version 2.17.3*.

Teichroeb, J. A. (2017). Methodological suggestions for inferring fear from vigilance. *Animal Sentience*, *2017.095*(Teichroeb on Beauchamp on Fear & Vigilance), 1–2.

Treves, A. (1998). The influence of group size and neighbors on vigilance in two species of arboreal monkeys. *Behaviour*, *135*(4), 453–481.

Treves, A. (1999). Within-group vigilance in red colobus and redtail monkeys. *American Journal of Primatology*, *48*(2), 113–126.

Treves, A., Drescher, A., & Ingrisano, N. (2001). Vigilance and aggregation in black howler monkeys (*Alouatta pigra*). *Behavioral Ecology and Sociobiology*, *50*(1), 90–95.

Treves, A., Drescher, A., & Snowdon, C. T. (2003). Maternal watchfulness in black howler monkeys (Alouatta pigra). *Ethology*, *109*(2), 135–146.

Unck, C. E., Waterman, J. M., Verburgt, L., & Bateman, P. W. (2009). Quantity versus quality: how does level of predation threat affect Cape ground squirrel vigilance? *Animal Behaviour*, *78*, 625–632.

van Shaik, C. ., & van Noordwijk, M. . (1989). The special role of male Cebus monkeys in predation avoidance and its effect on group composition. *Behavioral Ecology and Sociobiology*, *24*(5), 265–276.

Vasquez, R. A. (1997). Vigilance and social foraging in *Octodon degus* (*Rodentia: Octodontidae*) in central Chile. *Revista Chilena de Historia Natural*, *70*, 557–563.

Willems, E. P., & Hill, R. A. (2009). Predator-specific landscapes of fear and resource distribution: Effects on spatial range use. *Ecology*, *90*(2), 546–555.

Williams, S. T., Williams, K. S., Lewis, B. P., & Hill, R. A. (2017). Population dynamics and threats to an apex predator outside protected areas: Implications for carnivore management. *Royal Society Open Science*, *4*(161090).

Yaber, M. C., & Herrera, E. A. (1994). Vigilance, group size and social status in capybaras. *Animal Behaviour*, *48*, 1301–1307.

Table 1. Concordance correlation coefficient estimates for duration and frequency measures, across experienced and inexperienced and each definitional ethos. Asterisks indicate the strength of correlation estimates, *** indicates excellent correlation estimates (>0.9), ** indicates good (>0.75 and <0.9), * indicates moderate (>0.5 and <0.75), and no asterisks represents poor (<0.5).

	(1) Visual terminology CCC (95% CI)	(2) Head/eye movement CCC (95% CI)	(3) Non operationalised CCC (95% CI)	(4) Operationalised CCC (95% CI)	(v) 'Vigilance' CCC (95% CI)
Duration					
Experienced	0.939 (0.924,0.951)***	0.602 (0.538,0.659)*	0.899 (0.875,0.918)**	0.973 (0.966,0.978)***	
Experienced (excluding AA)	0.949 (0.936,0.960)***	0.538 (0.466,0.602)*	0.899 (0.874,0.919)**	0.967 (0.958,0.974)***	
Inexperienced	0.570 (0.454,0.666)*	0.241 (0.151,0.328)	0.452 (0.345,0.547)	0.943 (0.918,0.961)***	0.508 (0.392,0.609)*
Combined	0.740 (0.658,0.804)*	0.403 (0.305,0.493)	0.662 (0.571,0.737)*	0.965 (0.951,0.976)***	
Combined (excluding AA)	0.722 (0.636,0.790)*	0.366 (0.269,0.455)	0.634 (0.539,0.713)*	0.962 (0.946,0.974)***	
Frequency					
Experienced	0.763 (0.715,0.804)**	0.309 (0.242,0.374)	0.683 (0.625,0.734)*	0.880 (0.852,0.903)**	
Experienced (excluding AA)	0.767 (0.718,0.809)**	0.234 (0.165,0.302)	0.74 (0.686,0.786)*	0.882 (0.853,0.906)**	
Inexperienced	0.233 (0.130,0.331)	0.004 (-0.045,0.05)	0.204 (0.110,0.294)	0.848 (0.787,0.892)**	0.243 (0.137,0.343)
Combined	0.458 (0.354,0.552)	0.098 (0.05,0.15)	0.352 (0.249,0.447)	0.868 (0.818,0.905)**	
Combined (excluding AA)	0.432 (0.326,0.527)	0.06 (0.017,0.103)	0.330 (0.231,0.422)	0.863 (0.811,0.901)**	

Table 2. Summary of model results for duration response variable. Each column (i.e., AA, Obs 1 etc) represents each experienced observer, each row represents the fixed effects investigated within each observer's model. Cells are coded with asterisks according to the HDI-pd decision rule, * indicates that an effect has some uncertainty, ** indicate moderate evidence for an effect, and *** indicate strong evidence for an effect. Bold cells highlight where the pd was greater than 95%.

	AA	Obs 1	Obs 2	Obs 3	Obs 4
1) Visual terminology: "Animal is gazing or visually inspecting its surroundings"					
Intercept	-14.26 (-18.17, -10.44)	-13.31 (-17.07, -9.24)	-11.84 (-15.61, -8.05)	-12.7 (-16.6, -8.82)	-12.21 (-16.08, -8.12)
Engaged	0.09 (-0.02, 0.21)	**0.12 (0.01, 0.25)****	0.1 (-0.01, 0.21)*	**0.13 (0.01, 0.25)****	0.11 (-0.01, 0.23)*
No. Neighbours	-0.14 (-1.1, 0.79)	-0.08 (-1.02, 0.89)	-0.32 (-1.22, 0.61)	-0.01 (-0.99, 0.92)	-0.21 (-1.17, 0.76)
Neighbour Distance	-0.33 (-0.86, 0.18)	-0.25 (-0.77, 0.28)	-0.47 (-0.99, 0.02)*	-0.33 (-0.85, 0.2)	-0.25 (-0.78, 0.28)
Habitat	0.2 (-2.37, 2.86)	1.13 (-1.59, 3.7)	-0.05 (-2.58, 2.51)	0.49 (-2.17, 3.19)	0.61 (-2.09, 3.26)
Spatial Position	**-2.76 (-5.45, -0.04)****	**-2.99 (-5.72, -0.25)****	-2.46 (-4.99, 0.25)*	**-3.18 (-5.8, -0.36)****	**-2.9 (-5.56, -0.09)****
2) Head/eye movement: "Animal's head is up combined with side-to-side movement of the head and/or eyes"					
Intercept	-15.9 (-19.71, -12.23)	-20.22 (-23.62, -16.83)	-9.47 (-12.71, -6.05)	-14.48 (-18.43, -10.52)	-16.15 (-19.8, -12.18)
Engaged	**0.13 (0.02, 0.25)****	0.08 (-0.02, 0.18)	0.02 (-0.08, 0.12)	0.11 (0, 0.23)*	0.11 (0, 0.23)*
No. Neighbours	-0.16 (-1.06, 0.74)	-0.61 (-1.45, 0.2)	-0.64 (-1.44, 0.18)*	-0.08 (-1.03, 0.89)	-0.07 (-0.95, 0.88)
Neighbour Distance	-0.33 (-0.82, 0.18)	-0.25 (-0.72, 0.19)	-0.37 (-0.82, 0.07)*	-0.26 (-0.77, 0.29)	-0.15 (-0.66, 0.35)
Habitat	0.12 (-2.32, 2.69)	1.02 (-1.28, 3.36)	-0.37 (-2.72, 1.82)	0.34 (-2.32, 3.08)	1.62 (-0.88, 4.22)
Spatial Position	**-2.82 (-5.42, -0.24)****	-1.68 (-4.07, 0.65)	-1.13 (-3.48, 1.13)	**-2.97 (-5.8, -0.33)****	**-3.71 (-6.33, -1.02)*****
3) Non-operationalised looking/scanning: "Its head is up and eyes open"					
Intercept	-14.42 (-18.29, -10.48)	-15.26 (-19.07, -11.2)	-15.01 (-18.77, -11.25)	-12.98 (-16.92, -9.1)	-12.15 (-16.22, -8.34)
Engaged	0.09 (-0.03, 0.2)	0.1 (-0.02, 0.22)*	0.1 (-0.01, 0.21)*	0.12 (0, 0.23)*	**0.11 (0, 0.23)***
No. Neighbours	-0.14 (-1.1, 0.79)	0.06 (-0.9, 1.01)	0.05 (-0.87, 0.96)	-0.2 (-1.14, 0.76)	0.1 (-0.83, 1.08)
Neighbour Distance	-0.22 (-0.74, 0.3)	-0.37 (-0.91, 0.14)	-0.31 (-0.79, 0.21)	-0.34 (-0.88, 0.17)	-0.23 (-0.74, 0.3)
Habitat	0.94 (-1.76, 3.63)	0.86 (-1.74, 3.6)	0.78 (-1.81, 3.31)	1.47 (-1.18, 4.17)	0.54 (-2.24, 3.12)
Spatial Position	**-3.29 (-6.05, -0.55)****	-2.45 (-5.24, 0.25)*	-2.59 (-5.29, -0.01)**	**-2.84 (-5.49, -0.02)****	**-3.35 (-6.12, -0.62)****
4) Operationalised looking/scanning: "Its eyes are open, and its line of vision extends beyond its hands and the substrate, animal or object that they are in contact with"					
Intercept	-12.68 (-16.76, -8.94)	-12.59 (-16.4, -8.61)	-13.32 (-17.03, -9.5)	-13.21 (-17.1, -9.44)	-10.45 (-14.24, -6.44)
Engaged	**0.14 (0.03, 0.26)****	**0.13 (0.01, 0.25)****	0.1 (-0.01, 0.22)*	**0.12 (0.01, 0.24)****	**0.12 (0, 0.23)***
No. Neighbours	-0.09 (-1.04, 0.86)	-0.14 (-1.09, 0.81)	-0.03 (-0.93, 0.9)	-0.13 (-1.07, 0.8)	-0.29 (-1.23, 0.67)
Neighbour Distance	-0.29 (-0.81, 0.23)	-0.3 (-0.83, 0.21)	-0.31 (-0.82, 0.2)	-0.27 (-0.8, 0.23)	-0.36 (-0.86, 0.18)
Habitat	0.64 (-1.95, 3.38)	0.53 (-2.16, 3.14)	0.69 (-1.92, 3.28)	0.6 (-1.98, 3.27)	0.05 (-2.49, 2.74)
Spatial Position	**-3.05 (-5.68, -0.21)****	-2.62 (-5.4, 0.05)*	**-3.01 (-5.62, -0.35)****	**-2.77 (-5.51, -0.15)****	-2.34 (-5.01, 0.4)*

Table 3. Summary of model results for frequency response variable. Each column (i.e., AA, Obs 1 etc) represents each experienced observer, each row represents the fixed effects investigated within each observer's model. Cells are coded with asterisks according to the HDI-pd decision rule, * indicates that an effect has some uncertainty, ** indicate moderate evidence for an effect, and *** indicate strong evidence for an effect. Bold cells highlight where the pd was greater than 95%.

	AA	Obs 1	Obs 2	Obs 3	Obs 4
1) Visual terminology: "Animal is gazing or visually inspecting its surroundings"					
Intercept	-28.64 (-28.9, -28.39)	-29.08 (-29.34, -28.83)	-28.74 (-28.96, -28.52)	-28.96 (-29.22, -28.72)	-28.76 (-29.03, -28.49)
Engaged	0 (-0.01, 0)	0 (0, 0.01)	0 (-0.01, 0.01)	0 (-0.01, 0.01)	-0.01 (-0.01, 0)*
No. Neighbours	**-0.1 (-0.16, -0.04)*** **	**-0.07 (-0.13, -0.01)*** **	**-0.06 (-0.11, 0)* **	-0.05 (-0.11, 0.01)*	-0.04 (-0.1, 0.01)*
Neighbour Distance	**-0.03 (-0.07, -0.01)** **	-0.01 (-0.04, 0.02)	-0.01 (-0.04, 0.01)	-0.02 (-0.05, 0.01)	0 (-0.03, 0.03)
Habitat	-0.1 (-0.24, 0.07)	0.01 (-0.15, 0.17)	0.1 (-0.04, 0.24)	-0.07 (-0.23, 0.09)	-0.03 (-0.18, 0.12)
Spatial Position	-0.14 (-0.27, 0.08)*	0.05 (-0.12, 0.22)	-0.07 (-0.22, 0.07)	-0.01 (-0.17, 0.16)	0 (-0.15, 0.16)
2) Head/eye movement: "Animal's head is up combined with side-to-side movement of the head and/or eyes"					
Intercept	-28.52 (-28.79, -28.23)	-30.14 (-30.59, -29.69)	-29.23 (-29.48, -29.98)	-29.45 (-29.72, -29.16)	-29 (-29.3, -28.71)
Engaged	0 (-0.01, 0.01)	**0.02 (0, 0.03)* **	0 (-0.01, 0.01)	0 (-0.01, 0.01)	0 (-0.01, 0.01)
No. Neighbours	**-0.13 (-0.2, -0.06)*** **	-0.11 (-0.23, 0.01)*	-0.03 (-0.1, 0.03)	-0.01 (-0.08, 0.06)	-0.05 (-0.12, 0.02)
Neighbour Distance	**-0.09 (-0.12, -0.05)*** **	-0.05 (-0.1, 0.02)	0 (-0.03, 0.03)	0.01 (-0.03, 0.04)	**-0.05 (-0.09, -0.01)*** **
Habitat	-0.12 (-0.29, 0.06)	0.01 (-0.28, 0.3)	-0.07 (-0.25, 0.1)	-0.04 (-0.23, 0.15)	-0.02 (-0.21, 0.16)
Spatial Position	**-0.23 (-0.41, -0.03)** **	-0.08 (-0.38, 0.22)	-0.01 (-0.19, 0.17)	-0.07 (-0.26, 0.13)	-0.06 (-0.24, 0.14)
3) Non-operationalised looking/scanning: "It's head is up and eyes open"					
Intercept	-28.57 (-28.84, -28.3)	-28.79 (-29.04, -28.51)	-29.05 (-29.29, -28.82)	-28.84 (-29.09, -28.59)	-28.98 (-29.24, -28.74)
Engaged	0 (-0.01, 0)	0 (-0.01, 0)	0 (-0.01, 0)	-0.01 (-0.01, 0)*	-0.01 (-0.01, 0)*
No. Neighbours	**-0.1 (-0.16, -0.04)*** **	**-0.11 (-0.17, -0.04)*** **	-0.04 (-0.1, 0.02)	**-0.07 (-0.13, 0)* **	-0.04 (-0.1, 0.02)
Neighbour Distance	**-0.04 (-0.07, -0.01)*** **	**-0.05 (-0.08, -0.01)*** **	-0.01 (-0.04, 0.03)	**-0.04 (-0.07, -0.01)** **	-0.02 (-0.05, 0.01)
Habitat	-0.09 (-0.24, 0.07)	-0.02 (-0.19, 0.15)	0.1 (-0.06, 0.26)	-0.07 (-0.24, 0.09)	-0.03 (-0.19, 0.13)
Spatial Position	-0.1 (-0.27, 0.08)	-0.01 (-0.19, 0.17)	-0.07 (-0.23, 0.1)	-0.01 (-0.18, 0.17)	0.1 (-0.07, 0.27)
4) Operationalised looking/scanning: "Its eyes are open, and its line of vision extends beyond its hands and the substrate, animal or object that they are in contact with"					
Intercept	-28.35 (-28.61, -28.07)	-28.41 (-28.67, -28.14)	-28.7 (-28.94, -28.47)	-28.32 (-28.57, -28.06)	-28.76 (-28.99, -28.51)
Engaged	0 (-0.01, 0.01)	0 (-0.01, 0)	0 (-0.01, 0)	0 (-0.01, 0)	**-0.01 (-0.01, 0)* **
No. Neighbours	**-0.09 (-0.15, -0.04)*** **	**-0.08 (-0.13, -0.03)*** **	**-0.07 (-0.12, -0.01)*** **	**-0.12 (-0.18, -0.07)*** **	-0.04 (-0.09, 0.01)
Neighbour Distance	**-0.05 (-0.07, -0.02)*** **	**-0.03 (-0.06, 0)* **	-0.02 (-0.05, 0.01)	**-0.05 (-0.08, -0.02)*** **	0 (-0.03, 0.03)
Habitat	**-0.22 (-0.37, -0.08)*** **	**-0.16 (-0.31, -0.01)** **	0.03 (-0.12, 0.17)	**-0.19 (-0.34, -0.05)*** **	-0.02 (-0.16, 0.12)
Spatial Position	-0.03 (-0.17, 0.13)	-0.09 (-0.25, 0.07)	-0.04 (-0.19, 0.12)	-0.03 (-0.18, 0.12)	0.07 (-0.08, 0.22)