Insights into the applicability of existing online reports of mammal sightings as a viable source of conservation data

Elliot Tripp

A thesis submitted to the University of Gloucestershire in accordance with the requirements for the degree of MSc by Research in the school of Natural, Social, and Sport Sciences.

September 2023

Revised March 2024

Word Count: 31,390

Abstract

The global biodiversity crisis has accelerated the processes by which new and novel conservation approaches are hypothesised, tested, developed, and adopted. Nowhere is this clearer than in the developing philosophy of iEcology (internet ecology), which utilises vast amounts of digital content created for non-conservation purposes, and extracts conservation-relevant information from them.

Where this is predominantly completed through the automated collection and review of content from search engines, social media, and digital libraries, this study fills the niche of a manual approach. The processes by which iEcology extracts data from digital content often require financial investment, development of new technological skills, and/or collaboration that can often be prohibitive to researchers and sufficient survey coverage.

This study subsequently utilises a manual data collection approach which seeks to further increase the accessibility of wildlife data extracted from digital content. To evaluate this methodology, posts were located using the default search functions of Twitter, Instagram, Facebook, Flickr, WhatsApp, and the Latest Sightings app in line with ethical guidelines agreed upon by the University of Gloucestershire.

15,659 social media posts of mammal sightings in Kruger and Pilanesberg National Park were observed over a 114-day period based on the date the post was made. Specific, species-level data (species, count, behaviour, location) were recorded alongside observations about the post such as the content type, the type of account that uploaded it, discrepancies between sighting date and post date. These data were recorded to assess correlations between the parks, platforms, people, and the type of wildlife encounters they recorded.

An inventory of 55 mammal species was created by observing these posts. SACs were used to measure the effort taken to discover species on each platform, and a Bray-Curtis model was implemented to statistically express variances in community composition across platforms. A species abundance estimate was attempted using capture-recapture techniques but ultimately failed due to an inability to properly compensate for replicated sightings.

Additionally, focus groups were conducted among ecologically minded individuals at a wildlife reserve in South Africa, and a signal-coverage test was conducted at Pilanesberg National Park to determine the effects of internet accessibility on the uploading of a post about a mammal sighting. It was ultimately found that the preferences of individuals towards certain species and behaviours (both those determined by the focus groups and expected according to existing literature) and effort (time spent in a park) were more significant in the creation of a post than internet accessibility.

This study makes insights into these conclusions, the potential applications of these data, where this methodology may be improved, and its place alongside iEcology and automated collection processes.

Declaration

This thesis is the product of my own work and does not infringe on the ethical principles set out in the University of Gloucestershire's Handbook for Research Ethics. I agree that it may be made available for reference via any and all media by any and all means now known or developed in the future at the discretion of the university.

Any views expressed in this thesis are those of the author and in no way reflect those of the university.

Elliot M.D. Tripp

September 2023

Revisions submitted March 2024

DOI 10.46289/FCAR2332

Acknowledgements

For the opportunity to complete what is hopefully an inspiring, informative, and definitive work, I would like to thank Adam Hart and Anne Goodenough for offering me this studentship position. I hope that, despite the ups and downs, I have ultimately created a piece of work which we are all proud of.

Ruby Williamson, who I met at the University of Gloucestershire during our undergraduate studies five years ago, is the embodiment of my purpose and my commitment. She has encouraged a competition within myself where I strive to prove that I deserve to be loved by someone as kind and smart and brave and beautiful as she is, and I love her dearly.

Charlie Martin is the best friend one could ask for. Knowing him for less than a year at the time of submission, he has been a confidante, a source of constant inspiration and joy, and has dedicated hours of his afternoons and evenings to listening to me ramble about my data and asking him what word sounds smarter. I left a part of me in Cheltenham when I moved away, and he is that part.

To my parents and guardians, Kimberley Shearer, David Tripp, Barnaby Shearer, Caro Shearer, and Bob Shearer, I thank you for your constant support and interest in my studies, my hobbies, and my growth as a person. I hope I have made you all proud.

My nan, Maureen Clark, who is no longer with us, is who I believe I inherited my adventurous traits and my willingness to say yes to everything, despite not always being sure or qualified, from. She is still loved and greatly missed.

Table of Contents

List of Figures

List of Tables

Section 1: Introduction and Literature Review

1.1 Innovation in the Biodiversity Crisis

The current biodiversity crisis, characterised by widespread species extinction (Brooks, *et al*., 2009; Braje, 2016), habitat degradation (IPBES, 2019), and ecosystem (services) decline (Kasso and Balakrishnan, 2013; IPBES, 2018), has spurred the development and adoption of novel approaches to address these conservation challenges (Cernev and Fenner, 2020; Gray, 2018; Hák, *et al*., 2016; Nikolova, 2016). While it is not wholly accurate to say that this crisis has solely triggered these innovations, it has certainly accelerated the processes by which new conservation approaches are hypothesised, tested, developed, and adopted (Iacona, *et al*., 2019).

Several factors contribute to the biodiversity crisis, including habitat destruction (Holden, 2017), anthropogenically accelerated climate change (Lowe and Walker, 2015), pollution (Alexander, 2016), overexploitation of natural resources (Sbragaglia, *et al*., 2022), accelerated spread of transmissible diseases (Dwyer, *et* al., 2020), and the spread and dominance of invasive species in non-native habitats (IPBES, 2019). These threats have been intensified by human activities including deforestation (Rudel, *et al*., 2009; Wallach *et al*., 2015; 2018), urbanisation (Butchart, *et al*., 2019), industrial agriculture (Schötker and Wätzold, 2022; Smith, *et al*., 2015), expansion of infrastructure (Hornsey and Fielding, 2019), the exotic pets trade (Bush, *et al*., 2014), direct exploitation through harvesting for food and products (UNEP, 2021), and the refinement of petrochemical elements (Schötker and Wätzold, 2022; Smith, *et al*., 2015; EPA, 2022; Ehrenfeld, 2003). Due to the immense impact humanity has had on global ecosystems, a new geological epoch, which also seeks to define this 6th mass extinction event, has been proposed; the Anthropocene (Braje, 2016; Caro, *et al*., 2011). The validity and extent of the Anthropocene are still debated, with many arguing that it should be recorded alongside the onset of the industrial revolution (Steffen, *et al*., 2011), while others propose that the massive impact of humanity on the natural world could extend as far back as the earliest instances of agriculture (Gowdy and Krall, 2013).

In response to the biodiversity crisis within the Anthropocene, scientists and practitioners (such as conservationists and technologists) have increasingly turned to innovative technologies and approaches. By leveraging ever-developing information and communication technologies (ICTs) such as remote sensing (Vermeulen, *et al*., 2013), satellite imagery (Boyle, *et al*., 2014; Seeley and Asner, 2021), drones (O'Conner, *et al*., 2017), camera traps (Kucera and Barret, 2011), acoustic sensors (Browning, *et al*., 2017), and mobile applications (apps) (Jepson and Ladle, 2015); ICTs enable researchers to monitor biodiversity loss in real-time, collect large-scale data sets, and even engage citizens in conservation efforts (Mckinley, *et al*., 2017).

1.2 Data in Wildlife Conservation

Data play a vital role in combatting the biodiversity crisis. An understanding of species distribution (Chaterjee, 2023), threats, and temporal trends in wildlife populations is essential to developing effective conservation strategies (Field, *et al.*, 2005). An understanding of baseline conditions must be obtained before strategies can be developed, and this understanding must be underpinned by well-informed and comprehensive data (Kindsvater, *et al*., 2018). These data need to be accurate, useable, and to have ideally been collected over a long period of time (to improve robustness, sample size, and demonstrate temporal trends), and over suitable geographical scales (Goodenough and Hart, 2017). As significant as this preliminary understanding is, it is also important to note that the monitoring of a system after conservation interventions have been made is of equal or greater significance, both to assess success and facilitate adaptive management where necessary (Manley, *et al*., 2004). Ecological surveys should therefore collect comprehensive species-level data, ranging from fundamental presence/absence assessments to more detailed data crucial for understanding population dynamics (such as abundance and density) and ecosystem health (such as behaviour and distribution) (Beeton, *et al*., 2015).

At its core, presence/absence data serves as a binary indicator that a species is (or is not) extant in the surveyed area, allowing for the creation of species inventories. Beyond this, abundance, density, and distribution data provide critical insights into population and range trends (Manley, *et al*., 2004). These metrics can act as early indicators of changes to the ecosystem, such as habitat degradation or fragmentation (Llausás and Nogué, 2021). Moreover, behavioural observations offer valuable context into species interactions, resource utilisation, and responses to abiotic factors within the environment (Manley, *et al*., 2004). Measurements of age, sex, and biomass can serve as proxies for population health, and can highlight potential sources of threat within the system (Zijp, *et al*., 2017).

Presently, most ecological survey techniques require time-consuming and expensive excursions to conduct surveys or deploy equipment (Seidlitz, 2021). “Traditional” techniques, such as transects, camera trapping, and aerial surveys, can quickly become logistical and financial burdens that act as a barrier to effective conservation work (Goodenough and Hart, 2017; Popov, *et al*., 2021). Table 1 discusses some of the most commonly utilized ecological survey techniques. Due to time and budgetary constraints, these techniques are often not applied to the spatial or temporal extent that is needed to ensure effective conservation (Grantham, *et al*., 2009; dos Santos, *et al*., 2020) and the longevity of successful conservation efforts (Goodenough and Hart, 2017).

Method	Description	Pros	Cons	Literature
Walked Transects	The survey area is determined, often from preliminary remote study of the environment. An individual/group of individuals is deposited at the start of a course and begin walking. Participants stop walking when they see an animal and record data including time, location (often GPS), species, distance, angle, number by age, number by sex, total number, and details of the environment they are found in such as "near water" or "in burn block".	Simple method. Utilises common equipment. Thorough, when conducted with intent. Can be used as an environment to extract local knowledge if a guide is hired.	Humans are imperfect and will miss sightings – inaccuracy increases with distance from transect. Variability in detection by observers	Davis, 1982 Clancy, *et al.*, 1997 Goodenough, *et al.*, 2023
Driven Transect	Employing the same method as a walked transect but from a vehicle. A typical bucky layout will have a driver and passenger, then several participants arranged on modified seating in the bed of a truck. The truck will be stopped when a species is spotted and the same measurements as a walked transect will be taken. Odometer measurements can be taken as an easy means of gauging distance covered during a survey.	All of the same pros as walked transects. Vantage point from the elevated seating of the vehicle. Can cover a greater distance more easily.	Limited to driveable terrain. Less habituated animals will avoid the vehicles, decreasing chance of detection. Additional expenses in vehicle and fuel. Requires a member of the part to be able to drive in the survey country. Fewer opportunity to interact with driver/guide.	Seidlitz, *et al.*, 2021 Hart, *et al.*, 2022
Aerial Surveys (Manned)	Manned aerial surveys take place in a helicopter. A pilot navigates the survey area in a grid pattern and the crew conduct observations of the species. Often used for species counts over large areas, the aerial viewpoint allows for a greater distance to be covered and affords less opportunities for species to avoid detection. Veteran aerial surveyors and pilots can flush animals out of hiding to achieve accurate counts and have systems to avoid the recapture of individuals.	Effective. Fast. Pilots are often local or have even surveyed that area before instilling knowledge and professionalism.	Expensive. Requires the deployment of a trained professional who is likely not among the original composition of the survey team.	Pollock and Kendell, 1987 Caughley, 1977 Pollock, et al., 2006

Aerial Surveys (Unmanned)	With the development of drone technologies, unmanned aerial surveys are becoming more accessible and efficient. Drones equipped with cameras are flown to cover large areas in an aerial grid.	Most accessible means of conducting an aerial survey. Developing camera technologies are making detection and identification better. Records footage that can be reviewed by multiple people to ensure accurate identification and enumeration. Cameras can be changed for infrared to detect nocturnal species. Cover large areas of land. Relatively short learning curve. Drones are quieter than helicopters and other manned aircraft. Drones can be more easily transported internationally.	Drones have short battery lives compared to the fuel capacity of a helicopter. Remain expensive. Cameras provide a limited field of view that may lead to missed detections. Drones are fragile and a slight malfunction can cause catastrophic damage.	Vermeulen, *et al.*, 2013 O'Conner, *et al.*, 2017
Camera Trapping	Camera traps are typically placed in areas that a high frequency of passage by animals can be expected. They are triggered by sensing motion and instantly capture a (series of) image(s) of what triggered them. There is an extensive range of camera traps available on the market ranging from those intended for recreational use, to those for hunting, and models specifically designed for wildlife survey.	Variable selection of camera traps including features such as: Field of View. Range. Visible light/infra-red. View or image. Battery Life. Median model price $150. Traps can be baited with scent or food to encourage wildlife.	Collects large amounts of null data (false triggers). Less effective for smaller species which struggle to trigger the traps. Many traps needed to cover a large area quickly becoming expensive. Baited traps skew results.	Chapman, 1927 Rovero, *et al.*, 2008 O'Connell, *et al.*, 2011 Trolliet, *et al.*, 2014 Rowcliffe and Carbone, 2008 Trolle and Kery, 2005
Catch and Release	Most common in examining fish where large nets can be employed or invertebrates like moths where light traps are used, catch and release methods allow scientists to become	Allows closer, longer inspection of animals.	Stressful for the animals. Expensive to employ.	Merckx and Slade, 2014

	familiar with the species they are studying. For mammals, baited cage traps are often used.	Has led to the detection of many new invertebrate and fish species. Allows for assessment of animal health. Allows for attaching of a collar or chip to animal for tracking.	Unreliable due to dependence on many external factors. Covers a small area. Time consuming.	Jonason, *et al*., 2014
Tracking	In this method, animals are identified by their tracks and other physical markers such as lost feathers, evidence of grazing, and dung. It is possible to follow the tracks to the animal themselves but is not necessary to make preliminary assessments. Modern iterations of this method integrate GPS technologies to track animals and interpret behaviour based on migration patterns and interaction with other animals.	Simple method. Does not require specialist equipment. Encourages interaction and sharing of knowledge with locals.	Tracks often imperfect, disformed, or washed out leading to misidentification. Tracks are short-lasting. Requires specialist knowledge/ experience to be effective.	Kays, *et al*., 2015 Costa-Pereira, *et al*., 2022
Nocturnal Game Drives	Nocturnal game drives rely on the eyeshine generated by the *tapetum lucidum* that many mammal species possess. The *tapetum lucidum* contributes to better night sight, meaning that species which possess it are often crepuscular or nocturnal. Night drives exploit this fact and operate in twilight hours or at night, using spotlights to shine through shrubberies and foliage to capture eye shine. Thermal imaging cameras are often also employed. Once detected, a vehicle will stop and dedicate more of its lights to identifying the species. Night drives also give opportunity to detect nocturnal species, even if they do not have a *tapetum lucidum*, such as Aardvark.	Increased detection in areas of mid-density foliage. Quick. Engaging, leading to easy conscription of aid. Safer than nocturnal walked transects. In hot and arid environments, there is increased activity at dawn and dusk even among non-crepuscular species.	Detections often hard to identify as the animal flees the vehicle/light. Weather conditions can make staying out for long periods challenging. Limited by the terrain the vehicle can manage.	Hart, *et al*., 2022

The practical limitations ecological surveys experience persist despite a global rise in environmentalism (Pepper, 1996; Jeffries, 2005) and a social awareness of environmental issues (Ladle, *et al*., 2016), which has led to an increase in public fundraising for conservation issues (Jeffries, 2005). As it stands, however, it is often the case that this is still only sufficient to protect either a single species (Zhuang, *et al*., 2022) or a small geographic area (Schötker and Wätzold, 2022).

Despite all existing ecological survey techniques having unique strengths and limitations; it is the consistent limitation of time and money that most commonly leads to conservation efforts struggling (Bottrill, *et al*., 2009; Waldron, *et al*., 2013). These factors continue to spur innovation within the field of conservation (Brown, 2002), that can potentially reduce financial and temporal expenses in conservation projects and ensure that suitable spatial and temporal coverage can be achieved.

1.3 Wildlife Conservation in the Information Age

Technological advancements such as deep learning have already had significant impact on the means through which data are handled, analysed, and processed (Norouzzadeh, *et al*., 2018; Sbragaglia, *et al*., 2022). Many ecological survey techniques only target one species (Seidlitz, *et al*., 2021) with practices such as camera trapping producing large quantities of "bycatch" data (Smith, *et al*., 2017). Such footage is often dominated by abundant species in the area but can also include false triggers or instances where the image is not clear enough to identify a species (Ridout and Linkie, 2009). Processing this footage to extract ecological data is therefore often time-consuming, with little or no direct connection to the primary research and so is often overlooked. In some cases, "bycatch" data are added to large data libraries for other researchers to access after the fact.

Processing the quantity of images collected in such datasets is incredibly time consuming, and often relies on crowdsourced teams of volunteers to manually examine these images. Norouzzadeh, *et al*. (2018) found that deep learning systems can identify animal species with >93.8% accuracy, or 96.6% accuracy if it only identifies the content which it is confident about. These figures are comparable in accuracy to human labelling efforts, but for a dataset of 3.2 million images, deep learning systems reduced the time taken to identify the full dataset by 17,000 hours (8.4 years if assuming a 40-hour work week) (Norouzzadeh, *et al*., 2018). These technologies are ever developing and becoming rapidly more efficient, leading to even more significant time savings being possible (Liu, *et al*., 2024). Where climate change, habitat degradation, and decline of ecosystem services are occurring at their current rates, these massive time savings not only significantly reduce the expenditure of conservation efforts, but also help ensure that data are current and relevant (Leighton, *et al*., 2016; Dwivedi, *et al*., 2021; Sbragaglia, *et al*., 2022).

Alongside an increasing public interest in environmental matters and conservation efforts, there has been a growing utilisation of citizen science by conservationists (Willi, *et al*., 2019). Citizen science engages and empowers members of the public to actively participate in scientific research and monitoring efforts (Murphee, 2001; Petriello, *et al*., 2021; Powers, 2021). In response to community outreach, large datasets can be collected through citizen science (Dunn and Winkler, 1999; McShae, *et al*., 2023), just as large datasets can be manually examined by volunteers. Not only are these schemes advantageous to researchers who gain access to an expanded and eager workforce, but they are also beneficial to conservation efforts as community engagement can often be one of the most important factors to successful projects (Danielsen, *et al*., 2007; Denninger Snyder and Rentsch, 2020).

There are, however, additional considerations to be made from any crowdsourced data (Steger, *et al*., 2017). Public response to such projects is not guaranteed (Murphee, 2001), and therefore the submission of data is not either. It can also be challenging to backdate such data (Dunn and Winkler, 1999), unless referencing a past instance of citizen science, meaning that understanding temporal trends can take several years of ongoing survey effort; by which point, community engagement may have faltered (Barrow and Murphee, 2001; Hackle, 2001; König, *et al*., 2020).

To mitigate these issues of engagement, techniques that rely more heavily on passively sourced data are becoming ever more popular. iEcology (internet ecology) is one such practice. iEcology is a developing philosophy in ecology which utilises information from digital content generated for non-conservation purposes to extract data relevant to wildlife conservation (Jarić, *et al*. 2021; Pernat, *et al*., 2024). Online searches, social media platforms, and webpages generate unprecedented amounts of data (Hampton, *et al*., 2013, Giovos, *et al*., 2016). There are many ways that wildlife data can be extracted from these massive sources such as by (manually or automatically) classifying images and recordings (Falzon, *et al*., 2019), observing fluctuations in search terms (Anderson, *et al*., 2021), or tracking public engagement with conservation issues (Soriano-Redondo, *et al*., 2017; Johnston, *et al*., 2022).

iEcology studies can collect substantial amounts of specific species data with little effort. Panter and Amar (2022) used web-sourced photographs of Eurasian Sparrowhawk (*Accipiter nisus*) to study sexual dimorphism in the species by manually identifying the prey size within the photos. Studies like this often implement Application Programming Interfaces (APIs) to simplify and reduce the resource intensity of data collection (Di Minin, *et* al., 2015; Toivonen, *et al*., 2019). APIs (discussed further in section 3.2) return data to a researcher based on a series of terms they prescribe and can as such reduce data collection times drastically. Most digital spaces have some form of access to APIs, though their efficacy and reliability is somewhat inconsistent (Leighton, *et al*., 2016).

Of all of the common data sources in iEcology, social media platforms are perhaps some of the most diverse and abundant (Ghermandi and Sinclair, 2019). Many view the data collected from these platforms as passively crowdsourced data. Content on these platforms often relates to current affairs such as global events and public health (Lopriete, *et al*., 2021). As such, ecologists can often observe posts about, for example, invasive species (Encarnação, *et al*., 2021) that contain current and relevant data, or can mine posts to obtain retrospective ecological assessments (Hart, *et al*., 2018; Toivonen, *et al*., 2019).

Conservation culturomics, although not strictly under the same philosophy iEcology, studies fluctuations in online language usage to interpret trends in the natural world (Ladle, *et al*., 2016). By searching large bodies of digital text (typically search engine queries), public interest in topics can be tracked (Signorini, *et al*., 2011). Although culturomics tends to use this information to study the cultural saliency of particular species and issues, it can similarly be used to monitor environments (Galaz, *et al*., 2009). Proulx, *et al*. (2014), for example, found that searches for active ingredients in insect repellents correspond to seasons where mosquitoes are more present. Similarly, Pijanowski, *et al*. (2011) saw seasonal correlations in searches from the United States of America attempting to identify bird calls/insect noises/amphibian croaking, whereas searches for urban soundscapes did not vary seasonally.

As passively crowdsourced data, there are additional considerations to be made about the integrity of data collected by iEcology (Edwards, *et al*., 2021). Beyond several ethical considerations (which are discussed in section 3.2.1), understanding the motivation with which digital content is uploaded is also significant (Chua, *et al*., 2016; Arts, *et al*., 2015; Ghermandi and Sinclair, 2019). A combination of conservation culturomics and sociology may help to understand the motivations behind a social media user’s post and provide significant context, not only to the cultural significance of its contents (and potential avenues for conservationists to reach out to communities) (Mangachena and Pickering, 2021) but also for any potential biases, and subsequently sources of error (Kitchin, 2014; Ghaisani, *et al*., 2017; Hsiu-Chia, 2018). These potential biases could be either intrinsic (from the content creator’s personal preferences) or extrinsic (a result of a participant seeking external approval from the researcher and/or community). These motivations are seemingly underdiscussed in existing iEcology literature (Toivonen, *et al*., 2019; Jarić, *et al*., 2021).

There are conservation practices in digital spaces that do not fall under the banner of iEcology. By creating a purpose-built platform for opportunistic reporting of one-off observations, Sun, *et al*. (2021)

tracked American black bear (*Ursus americanus*) populations in New York state, USA, for 17% the cost of a typical catch and release programme for the same species. This mobile application (called iSeeMammals (iSeeMammals, 2024)) not only saved a substantial amount of money but was also significantly less labour intensive. By advertising the application to local students and hikers, the platform gained a sizeable userbase in a short period of time.

A similar application to iSeeMammals which was developed by volunteers and is free to use is the Latest Sightings mobile application (LSApp) (Latest Sightings, 2023, a). LSApp's primary foci are South African National Parks (SANParks) and game reserves, and it was developed specifically with Kruger National Park (KNP) in mind. The app was developed to support tourists and hobbyists by providing live, real time, and current data about wildlife sightings within this community, but has since collaborated with several conservation organizations, research projects, and even park rangers to achieve more holistic goals. LSApp is discussed further in section 2.7.

1.4 Aims, Objectives, and Relevance

This study aims to (1) determine how public social media posts relating to mammals vary across Twitter, Instagram, Facebook, Flickr, WhatsApp, and LSApp; and (2) determine what information held within such social media posts could be useful in an ecological and/or conservation context. Data will be collected from these platforms via manual retrieval of posts referencing mammal sightings in Kruger National Park (KNP) and Pilanesberg National Park (PNP), South Africa. KNP and PNP have been chosen due to their extraordinary richness of unique biodiversity, particularly of mammals, (IPBES, 2018), South Africa's historical reliance upon ecosystem services and natural resources to mitigate poverty in urban and peri-urban areas (Ward, *et al*., 2016; Kepe and Hall, 2020), and the level to which anthropogenic activity is contributing to biodiversity loss in the area (Xu, *et al*., 2016; UNEP, 2021). All of these factors mean that iEcology, in providing effective and rapid data collection for this area, may be one of the most important tools protecting South Africa's biodiversity in the Anthropocene.

The specific aims of this project are to:

(1) Determine how posts relating to mammal sightings in KNP and PNP vary over different social media platforms and apps.
(2) Establish what types of content these posts may contain.
(3) Critically evaluate what potentially useful data could be extracted from this content.
(4) Advise researchers on the potential benefits and applications of these types of data and the manual data collection approach.

These aims will be met via the following objectives, mapped to the original aims:

(1a) Locating mammal sighting posts from KNP and PNP using the built-in search and filter functions of the platforms.

(1b) Creating a dataset that records information from each post such as date of post, content type, and level of detail about the sighting included in the text description (where applicable).

(1c) Understanding how user behaviours vary across platforms and influence the content being posted by (1) reviewing existing understandings of social media user habits (Fan and Gordon, 2014; Parwani, *et al*., 2019) and (2) conducting focus groups among volunteers, staff, and students on fieldwork at a game reserve near PNP.

(2a) Recording specific data such as species, quantity, behaviour, location, and sighting time across all platforms.

(2b) Comparing the distribution of specific species and behaviours across platforms, with specific interest in the differences between LSApp (a dedicated platform) and the other social media platforms (where this species data are consequential of other user behaviours).

(3a) Create species inventories across platforms and use Bray-Curtis' measures, PERMANOVAs, and other statistical analyses (Chi-Square test for statistical association) and graphical representations to test the quality of data.

(4a) Draw critical conclusions about the successes and shortcomings of manual data retrieval techniques, and of the quality of data obtained from these platforms.

Section 2: Social Media Platforms

2.1 Social Media

Social media platforms such as X (formerly and henceforth referred to as Twitter), Instagram, Facebook, and WhatsApp encourage the distribution and exchange of image, video, and text content online. A large number of posts pertain to details of an individual's personal life including their recreation and leisure activities, personal life updates, and their vacations which, in the case of a safari holiday, often contain mammal sightings.

As the world recovers from the Covid-19 pandemic, international travel is once again rising and leisure travel, including wildlife-based tourism, is likely to continue to experience the growth it was before the pandemic (UNTWO, 2019). With the popularity of mobile phone technology, the majority of tourists have on their person a device capable of creating text, image, and video content, connecting to the internet to upload said content, and which are actively designed to promote social media use (Salehan and Negahban, 2013). Every social media platform presents a different suite of features and userbases which ultimately impacts the types and subject matter of the content uploaded to them. This creates a diverse and interesting range of conditions across these platforms that ultimately leads to diversity of content. Here, each of the platforms in this study will have its userbase and features briefly discussed to contextualise why they were chosen and why they may present content in the manner that they do.

Social media platforms change their features, userbase, preferred content, terminology, and even names over time corresponding with the preferences of advertisers on the platforms, legislation regarding online spaces and privacy, user habits, and the whims of the owners. However, there are some basic similarities in use and features and the language employed to describe them.

Typically, in social media, the term "post" is used as both a verb and a noun. The act to (make a) post is to upload content to a platform. This content then becomes *a* post. If a user created this content, they are the Original Creator (OC) and they have produced original content (oc). OC and oc are used almost interchangeably in online spaces, as both are indicators of the originality and ownership of content. Henceforth, this study will refer to the unique uploading of oc regarding a wildlife sighting simply as a "sighting", while a "post" in this study will refer to any piece of content collected as data including those that were the result of reposts and/or multiposts (defined below), and regardless of the type of content uploaded. The verb to post will be used to signify that a piece of content has been uploaded, created, redistributed, or in any other way propagated to a platform and may be attached to the process of creating a unique sighting.

While reposting and multiposting might often be treated as synonymous, an important distinction will be made here. A “repost” will be defined as a piece of oc taken from one platform and propagated to another. A “multipost”, however, will be defined as an instance where multiple OCs upload content about the same sighting, often to the same platform. Both reposts and multiposts involve the pseudo-replication of information, but a distinction is made here since: (1) in multiposts the content of the posts might differ on the perspectives of the OCs; and (2) multiposts can subsequently be reposted, which will exponentially inflate figures. This differentiation theoretically allows for the consideration of recapture rates within this study.

On most social media platforms, posts (whether they are oc or a multipost or a repost) are, by default, presented on a user’s timeline. Timelines are an order of presenting content to users that is primarily weighted by how recently they were uploaded and is the main space that users will observe when using a platform. Interaction with posts by other users impacts the order that they are presented on people’s timelines, with more interaction raising the prominence of the posts. This interaction is referred to as the platform’s algorithm (Balaji, *et al*., 2021) which is, in its simplest form, the extent to which a platform weights recency and interaction rates of a post when propagating it to a timeline.

Interactions come in two main forms. Firstly, “likes”, often called reactions, are a simple means for a user to engage with content they enjoy. Liking/reacting to a piece of content allows the OC and algorithm to know that a user enjoyed it, allowing for similar content (typically determined by the mutual reaction of users between two pieces of content) to be propagated to a timeline. On some platforms, simply liking a post will propagate it to that user’s followers within their timeline.

The second form of engagement is “comments”. The comments section attached to a post acts as a venue for followers to express their feeling about a post in text, rather than the symbols (emojis) that form a reaction. Comments can be reacted and replied to just as a post, acting as a forum for user engagement.

2.2 The Social Media Platform formerly known as Twitter

Twitter is a microblogging platform which launched in 2007 and built its identity on short content, with text posts being limited to 140 characters. This limit expanded to 280 in 2017 following a change in user habits on the platform. more recently, due to the addition of paid features to the application by the platform’s new owner in 2022, Twitter Blue (paid) subscribers can post up to 4,000 characters in every message. Another feature of the recent takeover is an attempt to rebrand the platform to X but, as many still refer to the platform as Twitter, that nomenclature will be followed in this thesis.

Despite much of Twitter's identity being built upon the text character limit, an observation of the average timeline will be mainly dominated by image or video posts (usually accompanied by explanatory or commentative text). All presentations of short content have found a home on Twitter where the concentration of current affairs content, engaged with debate beneath posts in "threads", and an algorithm that pushes highly engaged with content, in turn leading to higher engagement, in turn leading to greater promotion, has made Twitter the *de facto* soap box of the internet.

Twitter possesses several unique terminologies for its otherwise generic features. For example, a post is referred to as a "tweet", and the timeline as the "feed". The sharing of a post is referred to as a "retweet" and can either be a quick share that simply reposts the content to be seen by a user's followers, or made using the "quote tweet" feature which allows the user to add their own text description or comments to the post in addition to the shared content. The latter is often used for content that the user wishes to engage with or share a reaction to, while the other is used to share content that a user likes and/or thinks their followers would appreciate.

Search functions on Twitter are fairly nuanced. The infamous #Hashtag feature is a tagging practice that allows for better promotion and relevancy of content. Within the text of a post, often at the end or integrated throughout a post, users can add a "#" before key terms such as #subject matter, a particular #location, or a trending #topic (Scanfeld, *et al*., 2010). Deviation in the presentation of hashtags can lead to incomplete search results, this essentially making a post more difficult to find. For example, "#Country Name" may (seemingly inexplicably) return fewer results than "#countryname", simply because more users have adopted the latter due to the informal nature of language online. Beyond this, however, Twitter's search functions are quite simple. A search is by default sorted by Top (most engaged with recent content) but can be changed to Latest (most recently uploaded), People (accounts rather than posts) or Media (posts which contain the key word and have an image or video attached). The only filters that Twitter affords are whether the results are returned from public users or just followers, and location.

Twitter's algorithm favours current content. As such, it rarely promotes older posts to a user's timeline, leading to them receiving less engagement and therefore less promotion. Theoretically, for the sharing of real-time wildlife sightings, this is ideal since individual animals can move quite frequently and recency is subsequently key for those hoping to recreate the sighting.

2.3 Instagram

Instagram (etymology: “instant (camera) + “telegram”) is an image sharing platform that launched in 2010. It built its identity on the streamlined and visual sharing of memories and “the moment” by its users. Although users have the option to add text descriptions beneath their posts, the image or short video consumes the majority of a user’s screen and timeline, making any attached text secondary. The engagement of one user with a post on Instagram does not directly propagate it to their followers, and there are no options on Instagram to share posts to a timeline. Instead, there are options to send it as a direct message to another user, or to upload it to a story. Stories on Instagram (and Facebook – section 2.4) are posts that disappear after 24 hours and are presented in a dedicated area of the platform separate to the timeline. Beyond this, Instagram offers fairly standard social media features including likes and comments. The main quirk of the platform is the inability to post exclusively text posts on the image-dominated platform.

On Instagram, a tag serves as a means for a user to redirect viewers of their content to other relevant content. When a post is uploaded, an OC has two options (which are not mutually exclusive or dependent) to tag (which is different to #hashtagging) such content. The first option is the location. This is selected from a list generated by Instagram and is typically either (a) a broad description such as a city, region, or country; or (b) a specific venue. The inability of users to tag unlisted locations means that there are often instances where OCs may default to using a broad location if the specific venue they are attending is not listed, reducing the accuracy of location tagging on the platform. The second tagging option is that of accounts. When viewing a post with an account tagged in the image, a semi-transparent icon will be visible in the bottom left corner of the post. Interacting with this icon reveals a speech bubble-like overlay on the image with the tagged account and clicking/tapping on it will redirect a user to that account’s page. This feature is often utilised in group photos or posts with sponsored or collaborative content. Tagging an account in this manner serves a secondary purpose. If an OC’s (A) account’s privacy is set to public or another user (B) is a mutual follower of both A and tagged the account (C), B will receive a notification about a post relating to C uploaded by A, although the post will not be automatically propagated to B via their timeline.

2.4 Facebook

Facebook is the most popular social media platform, with the largest number of concurrent users and the greatest amount of content uploaded to, and consumed from, it. The platform launched in 2004 as TheFacebook, shortening to simply Facebook in 2005. The company that owns Facebook, Meta, also owns Instagram and WhatsApp. Despite this, the platforms have managed to retain their independent identities by retaining their unique features. While Instagram would best be described as a media sharing platform, Facebook is a true social networking platform. Followers on Facebook are referred to as "Friends" and although there are levels of privacy mediation that users may take, friends are typically mutual followers of one another. A user's timeline on Facebook will be populated by posts from one of two origins: friends and pages. Pages are non-personal accounts on Facebook and posts will enter a user's timeline when they follow said page or when Facebook's algorithms determine that a user may enjoy the page's content based on what said user interacts with on the platform. Facebook utilises a tagging system similar to Instagram, but this is predominately used to tag individuals rather than locations. Similarly, Facebook has introduced a disappearing Story feature, though this is more broadly adopted by pages and businesses than individual users.

The main appeal of Facebook for many is the presence of groups. Groups are communities made up of non-friend accounts who create and share content about a specific topic. The groups are semi-closed and moderated to ensure content remains pertinent to the intended audience. Thus, they have the potential to act as concentrated pools of relevant data for any conservation study.

2.5 Flickr

Flickr, which debuted in 2004, operates in a similar manner to Instagram but with many of the social aspects having never been introduced. Although it is possible to follow a profile, favourite (like) posts, and comment on content, there is no traditional timeline on the platform and nor are there any options to share or communicate via direct messages, as there often are on other platforms. Flickr abstains from these features due to its indented functionality and audience: the propagation of high-quality photography among hobbyists. Where Instagram seeks to capture the moment and is designed to be fast and snappy, Flickr has no such intention.

Instead, Flickr has a focus on the process, allowing and encouraging users to present information about their camera setup and the image-taking experience alongside their posts. Text descriptions are utilised differently on Flickr and predominantly share details about the captured event, rather than the narrative aspects or context. This practice is supported by features such as an additional "taken on" date alongside the upload/post date, and a tagging system unlike those of any other platform. Tags on Flickr can act as secondary text descriptions about specific details within posts. Due to the focus on image quality, there is a lower level of (file size) compression on Flickr that leads to a user being better able to enlarge features of an image with zoom functions. This combined with the tagging system OCs users to select features in the background of a photograph and attach a description to it, allowing viewers to zoom into the detail and appreciate it. The level of file size compression on other platforms reduces the quality of the images and would make such a feature less effective.

2.6 WhatsApp

WhatsApp is a messaging service (owned by Meta), opposed to a traditional social media platform. This means that there are no public likes, no followers, no timelines, and none of the other traditional features discussed above. However, due to its online hosting and end-to-end encryption, WhatsApp has not only superseded most of its messaging service competitors but has also become a favourite for closed groups similar to those present on Facebook. WhatsApp Groups (groupchats) are often used as an alternative to groups on other platforms with admins (moderators) from the community managing the members and content to remain safe and relevant. Latest Sightings promote their WhatsApp groups as a place to undertake the social networking aspects of social media that LSApp does not provide and use it as a site for their announcements and daily summaries.

2.7 The Latest Sightings Mobile Application

The Latest Sightings app (LSApp) was constructed for the specific purpose of allowing users to share wildlife sightings with fellow tourists in South African National Parks (SANParks). The LSApp was developed in 2011 and has many characteristics and features typical of social media platforms, although it applies them differently.

LSApp calls sightings uploaded to its platform "tings". Tings are encouraged to follow a structure that includes time, species, a description of the sighting, location (usually given via a combination of road name and direction/distance from the nearest road junction, similar to informal directions), behavioural observations, a visibility score (added part way through this project's data collection), OC's name, and a photo/video (optional) of the sighting. When a ting is made, it is uploaded to two parts of LSApp. The first is a timeline reminiscent of those present on all social media platforms. Users have the opportunity to "like" tings, though unlike other platforms this does not affect their propagation on the platform, and the timeline is purely chronological. Tings are also simultaneously uploaded to a ting map. This map uses the location services of the mobile device (rather than the directions provided in the post) to place the ting marker which, combined with the location information provided by the OC, can be used by other app users to recreate the sighting. Tings remain on the map for 24 hours by default (though a new paid feature allows users to see a 7 day history of tings), and are colour coded in relation to the time that has elapsed from the sighting being recorded and the time a user is looking at the map.

It appears that the Latest Sightings team takes tings from LSApp and propagates them to other platforms, primarily Twitter and WhatsApp. This process seems to be semi-automated, pending a manual review of content. This observation was made due to tings from before business hours, after park closures, and during loadshedding were often reposted after a longer wait than tings made when the Latest Sightings team would be active and possessed the infrastructure to complete the review.

Section 3: Methods and Ethics

3.1 Study Area

Amid a general movement towards the public and private management of natural resources rather than governmental ownership (Fabricius, 2013; Turner, 2013), a government wildlife park was forged. This park later became known as Kruger National Park (KNP) and is now one of the largest game reserves in Africa. KNP is a relatively thin strip of land in eastern South Africa, bordering Mozambique and Zimbabwe and covering 19,485km^2 of varied vegetation from open veld to acacia-heavy bush (Britannica, 2023). KNP was founded after observations were made by hunters of the declining populations of wildlife in the area. These *ad hoc* observations led to the area being protected for restricted hunting, and later converted into a National Park, now overseen by SANParks (Turner, 2013). Pilanesberg Game Reserve (a provincial National Park and henceforth PNP and discussed as though a National Park) covers just 572km^2 of South Africa's Western province, but experiences much greater variations in topography than KNP with elevation variations of 300-600 metres throughout the park. Figure 1 presents the locations of the park in relation to one another and South Africa.

KNP and PNP were chosen as the study areas due to their prevalence on LSApp which serves both parks alongside Masai Mara National Park, Lower Zambezi National Park, Kgalagadi Transfrontier National Park, and Marloth Park. Both parks support a wide array of large mammal species. These include the Big Five; black rhinoceros (*Diceros bicornis*), African elephant (*Loxodonta Africana)*, cape buffalo (*Syncerus caffer*), leopard (*Panthera pardus*), and African lion (*Panthera leo*), as well as a large assembly of other species such as zebra (*Equus zebra zebra*), giraffe (*Giraffa camelopardalis*), hippopotamus (*Hippopotamus amphibia*), impala (*Aepyceros melampus*), and blue wildebeest (*Connochaetes taurinus*).

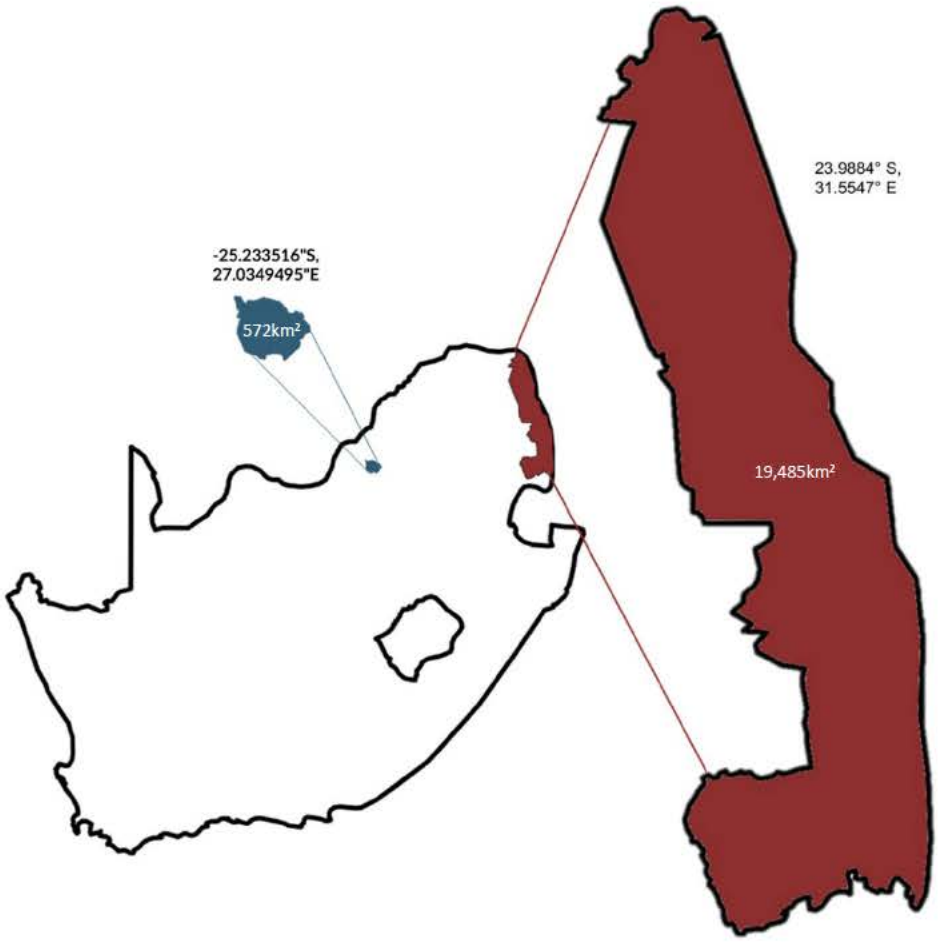

Figure 2: Locations of Pilanesberg (blue) and Kruger (red) National Parks scaled relative to one another. GPS coordinates are provided for each.

3.2 Data Collection Methodology

The data collection conducted in this project relied largely on manual data collection andinput into alphanumerically coded groups, rather than using APIs or external software as has been used, for example, by Hart, *et al*. (2018); Toivonen, *et al*. (2019); and Edwards, *et al*. (2022). APIs were excluded from the methodology due to the costs associated with them. APIs allow users to set data requests that the platform's servers automatically and periodically respond to (Leighton, *et al*., 2016). Twitter once had these openly accessible for free, but recently placed them behind a pay wall, supposedly to mitigate the presence of third-party bots on the platform (Kearney, 2019). Instagram and Facebook utilise Meta's graph API feature, but this was not comprehensive due to the nature of the platform's content and the limitations placed on publishing the resultant content and programmes. Flickr's API was free, but a combination of a small userbase and small development team meant that the API was underdeveloped (Bausch and Bumgardner, 2006). Utilising APIs for these platforms allows for the implementation of other technologies such as image recognition that would not otherwise be available on them, however, the cost is often prohibitive.

Although combining APIs and image recognition software could have returned a large number of posts rapidly and with reduced effort, they are often expensive on top of the cost of APIs, and their reliability is still questioned as they often fail to consider greater details and context; essentially having captured false positives (Edwards, *et al*., 2022). This study sought to achieve the same coverage as these processes but without the investment in APIs and software beyond those which would be considered standard for any researcher (Windows operating system, a web browser, Microsoft Office). Furthermore, even though AI will soon (within the next couple of decades) optimise the data collection and analysis process (Elish and Boys, 2018), making manual methods obsolete, this manual approach remains accessible and realistic for all to implement. The most challenging aspects of implementing manual retrieval were (1) implementing an efficient way of recording retrieved data (which will likely differ from person to person) and (2) ensuring accurate identification of the content being recorded (species, behaviour, etc). Both of these skills are ones that should be in the repertoire of ecologists at all levels.

For this study, preexisting personal accounts with no affiliation with APIs or an organisation, and with no preexisting membership of any relevant groups or pages, were used on every platform monitored. Posts were retrieved using the default search functions of each platform to obtain mammal sightings posts. Multiple search terms were used throughout the study, and these were modified to fit the preferences of each platform, as demonstrated in Figure 2; section 3.12. The impact of variable search terms was itself investigated in this study, as discussed in section 3.12. Once the search had been completed, the extracted posts were analysed based on the content of the text/image/video of the post and any text description.

All posts included in this study were visible to the account(s) used without said account(s) engaging in any misleading, peacocking, or coercive behaviours. Peacocking is the act of an unannounced researcher acting in a manner that would encourage participants to alter their behaviours (to match those exhibited by the researcher) (Spicker, 2011). Furthermore, no part of this study interfered with the end-user agreements of any of the platforms studied. Essentially, this study observed social media posts as one might observe visitor engagement with animals at a zoo. A zoo would typically be considered by most to be a public space with restricted access. Similarly, social media platforms are widely considered to be public, despite not necessarily being advertised as such. Unlike a zoo, however, social media platforms have settings that allow users to make their account private, or only visible to their followers. This study respected these boundaries, and this perceived privacy by not asking nor enticing users to change their habits or privacy settings. Data collection was limited to observing posts that were from accounts with public settings (achieved through only using the default search features of the platforms), as their use is described in the user agreements. No third-party applications or software were used to augment search results, meaning that every post located and included is one that the account should have been, by the platforms' definition at least, allowed and entitled to see.

3.2.1 Ethical Considerations

These actions corresponded with the ethical clearance granted by the University of Gloucestershire. Ethical clearance was deemed necessary because the definitions of public and private spaces and content online are still widely debated in academic circles and there were as such concerns that the study methods could be described as covert research (Spicker, 2011; Roulet, *et al*., 2017).

Ethical clearance was granted (see Appendix C) on the basis that the expected boundaries and privacy in online spaces described above would be respected. Users familiar with the platforms that they are posting on should be aware that their account comes with implicit privacy settings which directly or indirectly interact with the search function. As such, by utilising only these search functions, the privacy and boundaries of all users should be respected. Similarly, when gaining access to semi-closed groups such as those on Facebook or WhatsApp, it was ensured that the conduct of this research did not impede the terms of agreement of those groups. Where authentication was necessary through the completion of questionnaires on Facebook, these were completed with transparency and honesty. This restriction was implemented to ensure that the research did not (1) mislead any members of that groups; or (2) entice members of that group to change their habits.

Furthermore, in accordance with the University of Gloucestershire (2022) ethical guidelines, the British Psychological Society (2021), Government Social Research (GSR, 2022) guidelines, and other academic establishments and studies (Ahmed, *et al*., 2017) robust methods for data anonymisation and storage were also prioritised. The alphanumerical coding, as discussed in section 3.4, of observed features of posts was not only conducted to shorten data input terms, but also serves to anonymise the details of a post in its entirety.

Because of the steps taken, reconstructing a post with enough detail to find the original post and subsequently identify its OC would be a substantial challenge. Most importantly, accounts were identified as one of three simple categories: a Business account (owned and operated by an individual or group who used the account to market their goods or services), Personal account (an account serving an individual for recreational or novel purposes), or Tour Companies (accounts operated by tour companies within KNP and PNP). By anonymising data in this way, and minimising identifying features, the study conducted itself within ethical clearance guidelines.

There were additional concerns about the safety of heavily trafficked taxa such as rhinoceros and pangolin, and the sensitive data pertaining to them. The same anonymisation of posts that protects OCs was extended to all species present. In order to further protect endangered species, the specific details of any sighting, especially location, was not recorded nor will it be discussed in any depth within this study.

3.3 Data Types: Dates and Times

There were several key dates associated with each post that were recorded by this project. The first was the Post Date. This was the date on which a post was made according to the platform. All platforms studied here present this information somewhere attached to the post. On Android devices and via web browsers, Instagram formatted these dates as "X days/ weeks/months ago", but mobile iOS devices presented comprehensive dd/mm/yyyy dates. For this reason, an Apple iOS device was used for the majority of Instagram collection, while an Android device or the web browser versions of the platforms was used for every other platform. It was concluded after a brief observation period that this changing of devices did not significantly influence the posts returned by the search functions of the platforms. The same account(s) were used across all devices to ensure that any content preferences determined by the platforms' algorithms did not alter search results across devices.

Secondly, Sighting Date was recorded, where possible. This information was a prerequisite of LSApp tings and was assumed to be equivalent to the "Taken On" date attached to Flickr posts. For other platforms, the sighting date was occasionally provided in the text description attached to the posts. Many image posts were photos of a camera's viewfinder or screen taken on a mobile phone for easier upload; these camera screen images often presented a date and time watermark, which was taken as the Sighting Date.

These dates were redefined as Post Day and Sighting Day, where the first instance of a sighting was Day 1, and each subsequent Post or Sighting Date thereafter was assigned an appropriate day count. This format was used to calculate Post Lag as (*Post Day- Sighting Day)*, with the result being the time taken between a sighting being made and its OC uploading it.

The original data collection period spanned around 160 days, with the onset determined by practical factors beyond the concern of this study. The approach to data collection at the beginning of this term was unrefined, lacked focus and without the standards later implemented to increase the integrity of the data by removing biases towards definitions of primary species and behaviour (discussed in 3.7 and 3.10 respectively). Towards the end of this period, focus on and commitment to data collection were again lost as data analyses started. For these reasons, the dataset was trimmed to 114 days, based on the date that the data were collected not the Post or Sighting date, to ensure consistent coverage and a high standard of data. Subsequently, the range of data collection was: 16th February 2023 to 9th June 2023.

Sighting Time was gathered in a similar manner to Sighting Date. It was a prerequisite of LSApp tings and is similarly provided on occasion in the text description of posts or on the watermark of camera screen images. Sighting Time was recorded in a 24-hour format of hh:mm, and was assumed to be the local time zone (GMT+2).

3.4 Data Types: Account Type

As this project aimed to assess potential variations in user behaviour for posts pertaining to mammal sightings between platforms, it was important to record the general characteristics of an account type. This was noted but, as per the details in section 3.2.1, no details pertaining to the identity of the OC were recorded. The three account types considered here were Business (BU), Personal (PR), and Tour Company (TC). These distinctions were made due to the potential influence that an OC's status might have on the content they post. For example, it was hypothesised that BU and TC accounts may post more or less content of particular subject matters due to ulterior motives of promotion or profit. All official Latest Sightings accounts were classed as BU accounts alongside those of hotels and travel companies. While Tour Companies would have arguably been a type of BU account, they were treated here as a distinct group because of the increased concentration of competition locally and their increased interaction with tourists potentially influencing what they post based on presumed tourist preferences.

PR accounts were those operated by an individual for their own recreation and enjoyment or for the sharing of life events and experiences with friends and family. This classification encompassed the majority of social media users, including those who might be classified as professional influencers. All social media users could be considered influencers due to their ability to shape opinions, behaviour, and trends through their content creation and engagement (Bakshy, *et al.*, 2011). For the remainder of this study, however, "influencer" will refer to an individual who utilizes their platform to primarily or exclusively advertise or sell goods and experiences through interpersonal influence marketing (Flanagin, 2017).

3.5 Data Types: Post Content Type

Four post types were recorded in this study: (1) text posts; (2) image posts; (3) short-form videos; and (4) long-form videos. Text posts were any post where there was no image or video attached. These are wholly absent from Instagram and Flickr due to the content restrictions of those platforms. The vast majority of text posts were descriptive and included those that followed the format of statements of the species, count, and a description of behaviour. There were also some text posts, though much less common, which applied a greater level of emphasis on narrative and personification of animal behaviours.

Images were any posts where the primary content was an image or a series thereof. Some platforms offer users the ability to upload multiple images within the same posts. These are referred to as collages or albums or carousels depending on the platforms but will henceforth be referred to as collections. For the purpose of this study, each image in a collection was treated as a unique sighting or post. In the event where the same sighting contributed multiple images to a collection, only the first instance was recorded, using subsequent images of the sighting to complete any information that proved inaccurate with greater context, such as count if later images showed that the group was larger than first assumed.

Posts which were presented in video format were divided into video and long-form video based on their duration. Any video 59.99 seconds or shorter was classed as a video, whereas any 60 seconds or longer were classed as a long-form video.

There were often image and video posts which were close-up shots of animals, particularly faces, which provided no greater context about the presence of other animals around them or their behaviours. Sightings and posts like these were categorised by their content type as normal, but as an ethological observation could not be confidently made, these portrait posts were classed as an alternative to behaviour information being present, as discussed further in section 3.10.

3.6 Data Types: Multiposts and Reposts

Similar to how recapturing and recording an individual multiple times would inflate population numbers in a real-world survey, additional and/or repeated posts of the same sighting would inflate population estimates obtained by this study. To mitigate this, posts which appeared to present the same sighting multiple times were divided into two categories; multiposts and reposts.

The first instance of any post was recorded as normal. Beyond this, posts that seemed to be repeat instances of the same sighting were identified as either a repost (the exact same content shared to another platform) or multiposts (multiple pieces of content regarding the same sighting). Reposts were recorded by stating what platform the sighting was first recorded on, which allowed both the number and path of reposts to be recorded. Where possible, post time was used to determine the sequence in which posts were uploaded to and subsequently propagated from platforms. Twitter, Instagram, and Facebook all presented posts that were less than one day old as [posted] X hours [ago], while LSApp and WhatsApp both had hh:mm timestamps for their posts, and Flickr did not provide any such detail. Otherwise the platform on which the post was first recorded was assumed to be the original.

Multiposts were determined in an identical manner to reposts, but with the added caveat that they were present on the same platform as each other. Processing both data types relied on the recognition of key features of a post such as the species sighted, landmarks or vehicles in the background of the post, or repeated details in the text descriptions attached to them. This process was imperfect as it relied on recognition, memory, and judgement. Posts were only deemed multiposts when there was a high level of confidence about the individual piece of content.

Under these definitions, it was possible that an individual sighting could generate multiple posts on the same platform (multiposts) which were then duplicated on other platforms (reposts) causing exponential growth in the representation of that species, hence why both were recorded and distinguished.

Repost and multipost numbers were used to quantify unique sightings separate from posts. This was done to allow for abundance estimates to be attempted, where the Lincoln-Peterson Index would be applied to each of the most prevalent taxa as:

$$n = (C * R)/Rm$$

Where n=population, C=sum of taxa quantity by sightings, R=sum of taxa quantity by posts, Rm=sightings count-posts count. The estimates obtained through this calculation were then compared to existing estimates of species population from past surveys in KNP (Roocroft, 2021; Wray, 2023).

3.7 Data Types: Species

This study focused on mammals and did not document reptiles, birds, invertebrates, or flora despite their abundance within the parks. This decision was made for several reasons. Firstly, there was concern that documenting all species would result in too many species categories, many of which would have had very few posts. Secondly, preliminary observations of LSApp showed a major bias in the frequency of mammals tinged, with the vast majority of tings on any given day concerning mammals. Furthermore, Probert, et al. (2022) discuss the rates of uncertainty and misidentification in ecological community science which were often exacerbated among bird species identification where there are many more possible sources for misidentification such as complexity in plumage between males vs females vs juvenile in many species (Marini, et al., 2013; 2015). Mammal species are often more distinct in their differences, are typically larger than birds leading to greater clarity, and have fewer mobility options prolonging the exposure of users to the animal allowing for more time to compile an accurate identification or take a clearer image/video. This combined with overall higher rates of familiarity of mammals would hopefully have increased the accuracy of identifications (particularly among text posts which could not be verified and/or corrected by the researcher). Furthermore, mammals in South Africa are subject to intense and direct exploitation (Caine, 2007; Gibson, *et al*., 2019; Minin, *et al*., 2015; Montesh, 2013; Save the Rhino, 2023; UNEP, 2021; Yam, *et al*., 2018) and have few available means of avoiding indirect harms such as habitat fragmentation through urbanisation (Xu, *et al*., 2015) or loss due to agriculture (Sirami, *et al*., 2013).

Species were recorded with abbreviated codes. The primary species was always considered to be that which was first addressed in the text description in instances where multiple species were captured. In the case where visual content was recorded with no description (or where only the primary species was mentioned but other species are visible in the content), the primary species was taken to be: (a) that which appeared more in focus within the photo; (b) that which was first seen in the video; (c) the species closest to the top of the frame (as the species in the middle would likely be the one which appeared most in focus and thus ranked more highly according to criterion a). This hierarchy was also applied to rank additional species as secondary/tertiary/quaternary and was implemented to remove potential unintended bias towards any species potentially being deemed “more desirable” or “popular” during data input.

The majority of data analysis conducted by this study (discussed in section 3.13) were applied to the ten most prevalent species (or species groups) (also discussed as taxa), determined once data collection had concluded. These were (in alphabetical order of species and then of group): cape buffalo (*Syncerus caffer*), cheetah (*Acinonyx jabatus*), elephant (*Loxodonta africana*), giraffe (*Giraffa camelopardalis*), leopard (*Panthera pardus*), lion (*Panthera leo*), spotted hyena (*Crocuta crocuta*), wild dog (*Lycaon pictus*), zebra (*Equus zebra*), and the species group of antelope. Within this last category were: blesbok (*Damaliscus pygargus*), blue wildebeest (*Connochaetes taurinus*), bushbuck (*Tragelaphus sylvaticus*), duiker (*Sylvicapra grimmia*), eland (*Taurotragus oryx*), greater kudu (*Tragelaphys strepsiceros*), impala (*Aepyceros melampus*), klipspringer (*Oreotragus oreotragus*), nyala

(*Tragelaphys angassi*), reedbuck (*Redunca arundinum*), sable (*Hippotragus niger*), springbok (*Antidorcas marsupialis*), steenbok (*Raphicerus campestris*), tsessebe (*Damaliscus lunatus*), and waterbuck (*Kobus ellipsiprymnus*). These species were grouped under the banner of Antelope which will be discussed here as a cohesive unit due to their similar ecological significance and a potential lack of distinguishing between individual species by OCs. A pragmatic decision has been made to discuss antelope as a singular "species", "species group", or "taxa" throughout this study. This decision was made to simplify comparisons between the species group of Antelope and the presence and ecological significance of the other individual species discussed. Antelope will be referred to as a species, species group, or taxa interchangeably throughout this study. All other uses of species or taxa are purposeful throughout. For example, when discussing data collection and table 5; Antelope are not grouped in this manner and thus use of species is explicitly correct.

Together, these 24 species accounted for 88.91% of all sightings and 90.15% of all posts collected by this study.

3.8 Data Types: Individual Count

A count of the number of individuals was made where possible for every recorded post. This was either done by manually counting the number of individuals in an image or video, or by using the count presented in the text description if it seemed reasonable based on the typical group sizes assumed for the reported species. For example, a sighting claiming to have seen 40 impala was considered more credible than one claiming to have seen 40 leopards. It was also assumed that text descriptions would round numbers (typically up) to the nearest multiple of 5 or 10 and subsequently manual counting was conducted where images or videos were provided.

3.9 Data Types: Additional Species

Additional species were recorded and counted the same as primary species but were not considered an additional sighting or post. The behaviours of additional species were not determined as they were often not captured in enough detail for this to be accurately assessed or they were active participants in the behaviour of the primary species.

3.10 Data Types: Behaviour

Ethological assessments were conducted based on observations of the post. Where possible, the text description of a post was not relied upon to determine behaviour to ensure a level of consistency in assessment. OCs would be prevalent to context that the observer of a sighting or post would not, and this context can play a major role in determining the behaviour of an animal during a sighting. Furthermore, identifying behaviour is particularly subjective and where personification of animal behaviours may occur, personal biases and impressions should be avoided. To achieve this, a strict set of definitions for behaviours were created to insist that assessments were consistent and fair. Table 2 presents the definitions of the 8 most popular behaviours which accounted for observations made for 91.42% of all posts.

In addition to the listed behaviours, some posts were classified as Portrait. Although not a behaviour, a portrait reflected on the motivation of an OC to record a sighting beyond the subject matter of the species, in a similar manner behaviour was hypothesized to inspire a recording. Portrait posts were those which captured a close up image or video of an animal, typically their face, and provided inadequate context to assess the animal's behaviour.

It is noted that animals express behaviours to different extents and at different times dependent on their activity budget. This varies with species and in turn with time of year and season, and is impacted by factors such as weather, stress levels, or when they last ate. Definitions of behaviour attempted to avoid species or diet-specific language that they might be applied in equal measure to relevant content.

Table 2: The definitions of behaviours analysed in this study, created to ensure consistency between classification of behaviours.

Behaviour	Code	Description
(Dirt) Bathing	**BA**	Animal (partly) submerged in water or mud to remove dirt or parasites from its body or fur. The possibility that an animal could simply be in the water for recreation or to shelter from heat was considered in each instance of these behaviours based on level of motion depicted in the image or video, the species present, and general context.
Drinking	**DR**	Animals at a water source, seemingly for the purpose of hydration opposed to bathing or rest.
Feeding	**FE**	Any food-oriented action such as grazing, browsing, stalking, or killing prey, feeding on a prey item, scavenging, or using a salt lick to obtain nutrients. In instances of predator-prey interaction, the prey animal was not considered an additional species and was instead treated as a prerequisite to the primary species observing this behaviour.
Interspecies Interaction	**INT**	Any non-aggressive interaction between two individuals of different species including mutualistic behaviour such as mutualistic or commensal behaviour such as phoresy.

Intraspecies Interaction	**ISI**	Interactions between two or more individuals of the same species including playing, allo-grooming, or the rearing/feeding of young.
Mating/Courting	**MA**	Animals in the act of mating, courting.
Locomotion	**OTM**	Animal(s) moving from one location to another. Many species observe conduct their grazing or scavenging behaviours while moving, which would be considered as Feeding. For a sighting to be considered Locomotion, there must be no apparent alternative motive for the behaviour taking place.
Resting/Stationary	**RS**	Animal appears to be resting or sleeping.

Although behaviour data were collected for all species, most analyses were only applied to lion (the most sighted and posted species), leopard (the most reposted species), and antelope (the most underrepresented taxa relative to population abundance) to determine whether behaviour contributed to these statuses by impacting post habits.

3.11 Data Types: Location

Location data were taken in an attempt to ascertain the distribution of animals within the parks. Location data did not have a consistent structure across platforms, but where possible was recorded in the format prescribed by the most persistent source of location data: LSApp. This format was predominantly descriptions of how to drive to the location to recreate sightings, containing distances from junctions, hides, dams, road names, and nearby landmarks to achieve this. This ultimately makes them subjective and introduces high levels of uncertainty. To attempt to mitigate this, location data were only recorded when the distance from the landmark stated in the post was <6km.

Even with this attempt at quality control, location data on non-LSApp platforms proved incredibly sparse and the applicability of what location data were achieved was questionable. As such, it was not analysed in any significant detail, but was used when exploring the effects of mobile phone signal coverage (see section 3.15).

3.12 Data Types: Search Terms

Due to the significant differences between platform content, engagement, and use-case; search terms performed differently across platforms. This meant that utilising generic search terms across all platforms was in many cases ineffective and in others, outright detrimental to the collation of data.

In response to this issue, specific search terms were used for each platform, taking advantage of the unique features of the search functions and communities that said platform hosted. Figure 2 presents the process through which search terms were modified to best suit each platform. The terms present in the 'Included' phase returned the majority of results for the mentioned platforms. The reduction in results based on the term in the 'Included' phase here is primarily due to the order of application of those terms; many posts would respond to two or more queries but for the example below, search terms were tested in that order.

LSApp and WhatsApp do not support a search function in the same way as social media platforms; posts were simply collected directly from the timeline by recording the most recent post when the app or group was opened, and moving backwards in time until the last post recorded by the previous session was reached. Flickr was excluded from this systematic review also, as there were too few relevant posts on the platform for a standardised posting format to have been adopted by the community. As such, generic search terms yielded the majority of results gained here, but modification of terms did not significantly improve the data collection experience.

3.12.1 Groups and Pages Accessed

Search terms were not the only metric that dictated search results on all platforms. For Facebook and WhatsApp, dedicated groups and pages were accessed to increase the concentration of relevant content and to access posts that would not necessarily be returned by searches otherwise. Accessing these groups was done in accordance with the boundaries discussed in section 3.2.

The WhatsApp group chats accessed were those expressly made for LSApp users who were visiting KNP and/or PNP, which are public and advertised on the Latest Sightings website and their other social media accounts. No other KNP or PNP sighting-specific WhatsApp groups could be found. It is possible that some of the Facebook groups that were not joined for this study (due to entry questionnaires requiring inappropriate conduct as discussed in section 3.2) would have also had independent WhatsApp groups, but none of the Facebook groups joined did.

Although they will not be named specifically here (to anonymise their members), the Facebook Groups accessed were chosen based on their relevance to the study (i.e. focussing on wildlife sightings in KNP or PNP). These groups were only joined and accessed if completing the entrance questionnaire would not rely upon the researcher conducting any deception or peacocking. The process of locating these Facebook groups followed a similar logical framework to the creation of platform-specific search terms. The 'Included' search terms were used, with results filtered by 'Group'. An attempt was made to join the group if it was deemed acceptable according to the boundaries discussed in section 3.2. If no relevant posts were found upon a review of the group, it was left (to reduce the volume of irrelevant posts encountered later in the study).

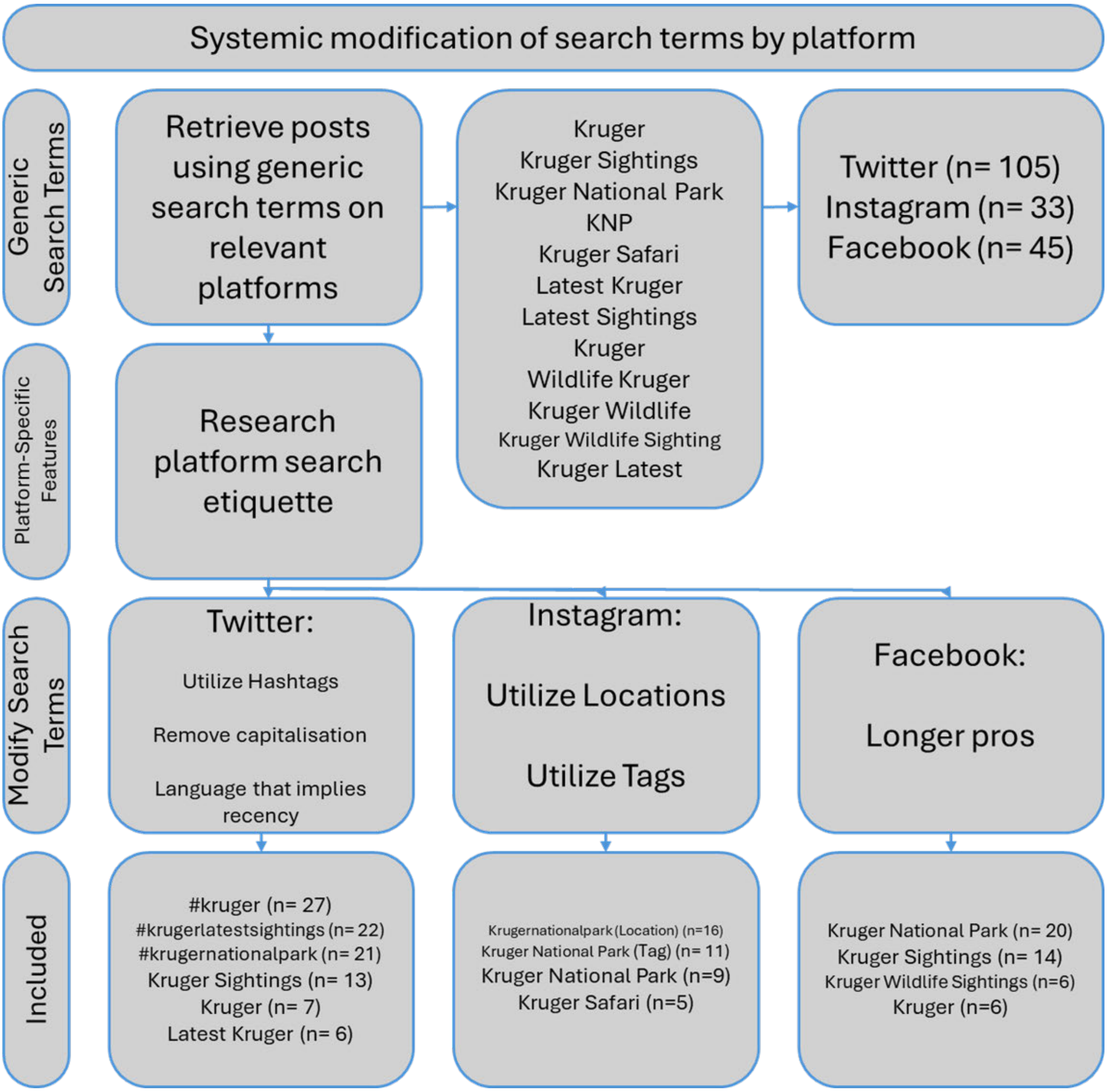

Figure 3:The systemic review and modification of the search terms used to collate data from platforms where search terms were the primary means of doing so. Model based on a PRISMA Flow Diagram (Page, et al., 2020). n=the number of posts found using these search terms over a 3-day test period retroactive to the main study, conducted to illustrate this process. For Generic Search Terms this was 29th January 2024-31st January 2024, and for the Included search terms this was 4th February 2024-6th February 2024, inclusive.

3.13 Analysis of Post Data

The data collected in this study was converted into two datasets. Dataset one included all sightings (unique recordings of mammal sightings being uploaded), while dataset two contained all posts (including all instances of multiposts and reposts). A fundamental understanding of the community representation in online spaces and across platforms was achieved by creating a summary species inventory for each platform. To further interpret the differences between the representation of communities across platforms, a Bray-Curtis measure of community dissimilarity was used supported by a pairwise one-way PERMANOVA (Permutational Multivariate Analysis of Variance) and Chi-squared test for statistical analysis to interpret the statistical significance of association between (1) the representation of individual taxa across platforms and (2) the representation of all taxa across individual platforms. The Bray-Curtis model and one-way PERMANOVA were completed utilising the PAST (PALeontological STatistics) software package (Hammer, *et al*., 2001). Additionally, Species Accumulation Curves (SACs) were constructed to represent the rates at which sightings on new platforms recorded new species for the first time; equivalent to the effort required (days spent) to discover new species on social media platforms and LSApp.

Alongside graphical representation of covariate analysis, Chi-squared tests for statistical association were applied to suitable matrices to determine whether a statistically significant association was present between two or variables. Due to the volume of data, number of parameters being recorded, and the interconnectedness of these parameters; data analysis was complex. Table 3 aims to illustrate the analytical framework used by posing a series of questions that could be asked of the data and then explaining what presentation and inferential tests were used to answer each question. These are not traditional "research questions" but part of a question-oriented analytical framework created to ensure each analysis was meaningful.

3.14 User Habits Focus Groups

During fieldwork to a wildlife reserve in the Northwest Province of South Africa (centred 25°15'S 27°17'E) in May 2023, focus groups were conducted among a sample of students, volunteers, and academics. Participants were not selectively chosen but rather comprised of individuals who were either on the same university fieldwork trip as the researcher, or whose stay or stint of volunteering coincided with said trip. Participants were asked several days in advance if they would be willing to complete these focus groups, and written consent was obtained immediately prior to their completion via a consent form that also informed participants of the project's research goals and how their answers would be used (Appendix C, Figure C2). This form was required as part of the ethics approval process for this study according to the University of Gloucestershire's guidelines (Appendix C, Figure C1). As the subject matter of these focus groups was non-contentious, and all participants were adults with the ability to provide informed consent, full ethical clearance for these groups was not deemed necessary.

There was a variation of experience level from those who were visiting South Africa for the first time and those who had been many times before. The majority of participants (89.2%) were white Europeans, with another 7.1% being individuals of dual nationality with South Africa, and the remaining 3.5% being New Zealanders who had emigrated to Europe. The modal age of participants was 19, and respondents were divided equally by sex, with only one participant having a gender identity different to their gender as assigned at birth. Individual focus group sizes ranged from 2 to 8 of these individuals, who were grouped based on their availability in relation to their other commitments at the reserve. As such, each focus group was essentially fully comprised of students (at undergraduate or master's level), or volunteers, or academics. Interviews spanned 7-15 minutes, though this did not correlate with group size but rather individual engagement with the questions. On the game drive in PNP (discussed in section 3.15), a volunteer of the wildlife reserve visited was engaged in informal conversation on the same topics as the focus groups throughout the day, though no recording was made. This volunteer was aware of the researcher's presence and goals in advance of this game drive. They verbally agreed that notes could be made about observations that occurred through these conversations and agreed to retroactively sign the same consent form as the focus group participants so that any significant points collected from this ad-hoc interview could be included in the study. Their participation was held to the same levels of anonymisation, data protection, and right-to-withdraw as the structured focus group participants.

Focus groups were held in an informal setting and recorded for later review, with these recordings being managed in a manner identical to the collected post data for the protection of privacy of all participants. To increase anonymity, attempts were made to avoid using names during recordings. Furthermore, there would be no need to identify any individual in the review of these focus groups.

*Table 3: The matrix of data analysis conducted in this study. *Specific named taxawere lion (the most prevalent species), leopard (the most reposted species), and antelope (the most underrepresented taxa relative to population abundance).*

Question	Presented as	Details
Do platforms vary in their frequency of mammal sighting reports and subsequent posts, irrespective of the park in question?	Graphical representation	Count of sightings and posts per platform (n=6). KNP and PNP pooled.
Do post rates vary by platform over time?	Graphical representation	Count of posts per day (n=114). KNP and PNP pooled.
Are all platforms equally likely to represent the presence of a species?	Summary table of independent species sightings by platform	Presence-absence of all observed species (n=55) by platform (n=6). KNP and PNP pooled.
Are species which are more popular (account for a higher percentage of posts) the result of more reposts?	Tabular representation of trends	Count of sightings of a species (n=10) compared to count of posts of a species, expressed as a ratio. KNP and PNP pooled.
Are species which are more frequently sighted (account for a higher percentage of sightings) subject to more reposts?	Tabular representation of trends	Count of sightings per species (n=10) compared to repost ratios thereof. KNP and PNP pooled.
Is there an association between platform and species community composition?	Chi-square test for statistical association of platform and species, Bray-Curtis's measure of community similarity, pairwise one-way PERMANOVA of species representation across platforms, graphical representation	Count of sightings per species (n=10) against platform (n=6). KNP and PNP pooled.
Do species that are frequently reposted contribute more to community dissimilarity?	Numerical representation in text	Repost ratio compared to Bray-Curtis' measure of community dissimilarity contribution. KNP and PNP pooled.
Does effort to detect (a) species vary by platform?	Graphical representation	Days (from n=1) until a species (n=10) was detected by a sighting on a platform. KNP and PNP pooled.
Does the capacity to represent species in sightings vary by platform?	Graphical representation	Count of first instance of species (n=55) detection per day (n=114). KNP and PNP pooled.
Is there an association between National Park and account type?	Chi-square test for statistical association, graphical representation	Count of sightings per account type (n=3) against National Park identification (n=2).

Is there an association between National Park and platform usage?	Chi-square test for statistical association, graphical representation	Count of sightings per platform (n=6) against National Park identification (n=2)
Is there an association between platforms and account type?	Chi-square test for statistical association, graphical representation	Count of sightings per account type (n=3) against platform (n=6). KNP and PNP pooled.
Is there an association between National Park and species community composition?	Chi-square test for statistical association	Count of sightings of a species (n=10) compared no National Park identification (n=2).
Is there an association between the content type of a post and the species it is recording?	Chi-square test for statistical association, graphical representation	Count of sightings per species (n=10) against content type (n=4). KNP and PNP pooled.
Is there an association between account type and species community composition?	Chi-square test for statistical association	Count of sightings or posts per species (n=10) against account type (n=3). KNP and PNP pooled.
Is there an association between account type and content type?	Chi-square test for statistical association	Count of sightings or posts per account type (n=3) against content type (n=4). KNP and PNP pooled.
Is there an association between content type and species, regardless of platform?	Chi-square test for statistical association, graphical representation	Count of sightings per species (n=10) against content type (n=4). KNP and PNP pooled.
Is there an association between content type and platform, regardless of species?	Chi-square test for statistical association, graphical representation	Count of sightings per content type (n=4) against platform (n=6). KNP and PNP pooled.
Is there an association between content type and platform, considering taxa?	Chi-square test for statistical association, graphical representation	Count of sightings of a species per content type (n=4) by platform (n=6). Tested for specific named taxa*, KNP and PNP pooled.
Is there an association between behaviour of a taxa and the platform to which a sighting is reported?	Chi-square test for statistical association, graphical representation	Count of sightings of a species per behaviour type (n=10) by platform (n=6). Tested for specific named taxa*, KNP and PNP pooled.
Is there an association between behaviour of a taxa and the content type through which a sighting was recorded?	Chi-square test for statistical association, graphical representation	Count of sightings of a species per behaviour type (n=10) by content type (n=4). Tested for specific named taxa*, KNP and PNP pooled.
Is there an association between species and its prevalence as portrait posts?	Chi-square test for statistical association, graphical representation	Count of portrait posts of a species (n=10) against platforms (n=6). KNP and PNP pooled.
Can sightings and posts be used to estimate species abundance?	Lincoln-Peterson mark-recapture index, graphical representation	n=(C*R)/Rm; n=population, C=sum of taxa quantity by sightings, R=sum of taxa quantity by posts, Rm=sightings count-posts count. KNP only.

Focus groups opened with a series of open-ended questions about participants' social media habits at home and while on holiday, including how frequently they would post, whether this rate would increase during holidays, and what forms of content they engaged with, propagated, or created. These opening questions led to a shift to informal conversation where questions were not necessarily asked, but points posed and engaged with by participants. These had no formal structure but typically followed the format of: (1) whether participants intended on using social media while at the game reserve; (2) whether participants intended to use social media on their rest day at PNP; (3) whether they had been able to uphold these intentions, and if not, what had prevented them from doing so; (4) whether they were engaging in a "digital detox"; (5a) what they would prioritise in terms of post content; (5b) if there were any particular species or events that they were actively seeking to see and subsequently post about; (6) where they would post and what their motivations behind a post were; (7) whether they had heard of LSApp and/or whether they would use it if they had; and (8) what steps participants would take to protect endangered species such as rhinoceros if and when posting about them on social media.

In order to ensure that participants had sufficient experience to answer these questions, focus groups were first conducted among participants who had visited the reserve or been on a safari holiday/excursion before, with less experienced participants being interviewed towards the end of their stays after they had been able to experience the environment themselves. The exception was one group of volunteers who were interviewed the day after their arrival to the reserve as this was towards the end of the researcher's excursion and otherwise this group would have been missed.

3.15 Signal Coverage and Latest Sightings Mobile Application Ting Distribution

On the same excursion to South Africa in May 2023, a field day was conducted at PNP in an attempt to test the accessibility of phone signal, which was substantial enough to upload wildlife sightings, primarily via LSApp. Four devices (three Android and one iOS) on multiple service providers (O2 (Android), EE (Android), and GiffGaff (Android) from UK devices, and a local Vodacom SIM (iOS device)) were used in the park during a day-long game drive. Tests to determine signal strength were conducted by attempting to load web pages, emails, or social media feeds. If this task could be completed and the device displayed a signal strength of H+ (3G) or greater, the device was considered to have connected. This test was conducted randomly/conveniently when the driver could safely stop to check their device, or when a sighting had brought the vehicle to a stop naturally. The number of devices that met the prescribed conditions was recorded at each location. It quickly became apparent that the likelihood of devices connecting was determined by their network and therefore the identity of each device was not recorded in instances where not all devices were connected. Instead, with one exception, the hierarchy displayed in Table 4 was representative of all records.

Table 4: The SIM IDs and Local Network Providers based on the number of successfully connected devices on the PNP game drive.

No. Devices Connected	SIM ID/Local Network Provider
1	Local Vodacom/Vodacom ISP
2	Local Vodacom/Vodacom ISP + O2/MTN Fibre ISP
3	Local Vodacom/Vodacom ISP + O2/MTN Fibre ISP + EE/MTN Fibre ISP
4	Local Vodacom/Vodacom ISP + O2/MTN Fibre ISP + EE/MTN Fibre ISP + GiffGaff/Telkom

To complement these data, the ting maps generated by LSApp were collected to create a map of signal coverage and ting distribution. Ting maps for PNP were collected from 50 consecutive end-of-day messages present on the corresponding WhatsApp group. These maps were created manually by the Latest Sightings team screenshotting and cropping the images from LSApp and as such were not standardised dimensions. Maps were resized using dams to ensure consistent scale, and the ting marked in the bottom of the icon provided (which were now inconsistent in size) on the new ting map overlay. This overlay was then presented on a map derived from the official road map of PNP provided by the park themselves, accompanied by the signal coverage data.

Section 4: Results

DISCLAIMER: The unorthodox colour palette for graphs in this section and throughout has been chosen to increase contrast between categories and minimise the use of reds and greens within the same figure (and subsequently distance them) to increase visibility for the researcher who experiences deuteranomaly; a common form of red-green colour blindness. Graphs with many categories may have coloured backgrounds to the same end. Appendix B will present copies of all figures with many categories with the default colour palette of Microsoft Excel.

4.1 Overview

Across the six platforms, 15,659 posts were taken as datapoints which arose from 8,483 unique sightings uploads. The distribution of these posts and sightings across platforms can be seen in figure 3. Sightings from LSApp accounted for the most of any platform observed and WhatsApp contributed the fewest. Twitter, however, returned more posts than LSApp, and WhatsApp posts accounted for more than Instagram, Facebook, and Flickr, with searches on Flickr returning the fewest posts.

There was an average of 137 posts per day, and a daily range of 51 to 324 posts per day across all platforms and both parks. Instagram and Facebook showed the greatest variation in post frequency relative to their abundance. Figure 4 presents how no single platform appeared to disproportionately influence changes in post frequency over time.

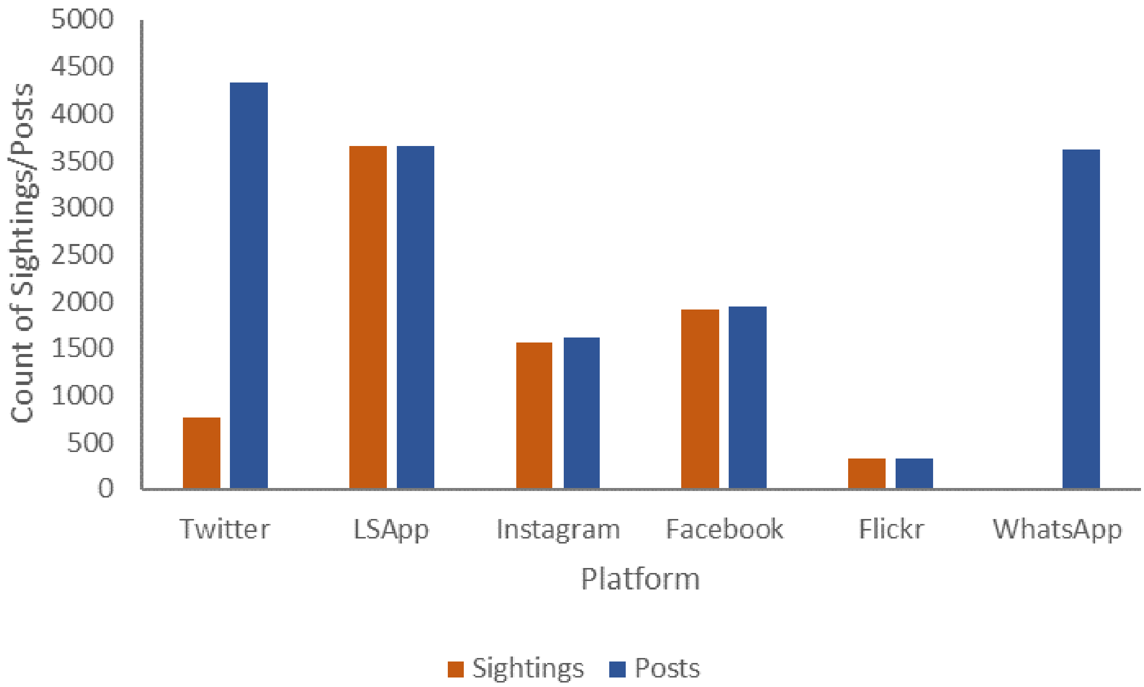

Figure 3: The distribution of sightings (unique instances of a mammal sighting being made, recorded, and uploaded) and posts (including reposts and multiposts) returned by searches on each platform over the 114-day study period.

The vast majority of posts regarded sightings made in KNP with 13,339 posts vs PNP's 2,320, which is approximately 470% more. KNP was the site of 7,274 sightings vs PNP's 1,209, approximately 500% more. This indicates different levels of user activity and propagation of posts within parks. Of the 55 unique species detected, 26 were recorded in PNP (47.27%) and 45 (81.82%) in KNP. Subsequently, if each of these 26 or 45 species were to receive an equal number of sightings from their parks, PNP would generate 89 sightings per species whereas KNP would generate 297 (approximately 230% more than PNP per species). KNP sightings detected 70% more species than sightings in PNP, therefore it would be expected, based on visitor numbers, that there were 410% more sightings posted from KNP than PNP, but in reality the value was closer to 477% additional sightings, indicating greater user activity at KNP.

As seen in figure 4, there were only two major peaks in post rate which occurred on and around day 53 (9^{th} April 2023) and day 73 (29^{th} April 2023), peaking at 324 and 306, respectively. Both peaks fell on weekends with day 53 falling within the Easter period and day 73 following South Africa Freedom Day, both of which are annually celebrated in South Africa. On the 29^{th} April 2023, there were three times as many posts about PNP sightings compared to the 9th April 2023, with 60 posts versus 19 posts. KNP-based sightings accounted for 246 and 305 posts on those same days, respectively.

4.1.1 Species Identified by Platform

A total of 55 unique mammal species from 25 taxonomic families were reported in the observed posts. The family with the most numerous species was Bovidae, with 17 species recorded, but individual lion and elephant sightings generated more posts than all Bovidae species combined. Table 5 lists the platforms where species presence was reported as sightings.

Only five species were reported on all six platforms: buffalo, leopard, lion, spotted hyena, and wild dog. Fifteen species were present on five platforms, with WhatsApp being the only platform where they were absent. LSApp, Facebook, and Flickr each recorded unique species not found on other platforms. LSApp included dassie (rock hyrax), elephant shrew, porcupine, scrub hare, side-striped jackal, red hartebeest, and Wahlberg's epauletted fruit bat. Facebook recorded blesbok, while Flickr sightings alone recorded the presence of forest shrew, little free-tailed bat, and springhare.

Overall, WhatsApp sightings achieved just 9.09% coverage. Twitter sightings achieved the second lowest at 56.36%. Facebook, (61.82%), Instagram (63.64%), and Flickr (65.45%) all covered around two-thirds of observed species. LSApp sightings covered the majority of species (81.82%). Observing post data presented substantially different trends. LSApp, Instagram, Facebook, and Flickr did not cover any more species, but Twitter's coverage rose to 89.09%, and WhatsApp to 81.82% (now covering all the same species as LSApp). This was due to the reposting of LSApp sightings to these platforms. The four species Twitter covered which LSApp did not were southern white rhinoceros, black rhinoceros, bushbuck, and tree squirrel.

Of the ten taxa which were most prevalent, defined in section 3.7, sightings and posts of lions were the most common, with a posts count of 4,827 and sightings of 2,040. Elephant were only reported half as much (2,449 posts; 1,071 sightings). Leopard were the third most prevalent species at 2,406 posts generated about 884 sightings. The least posted and sighted of the ten most prevalent species was zebra at 506 and 344 instances respectively.

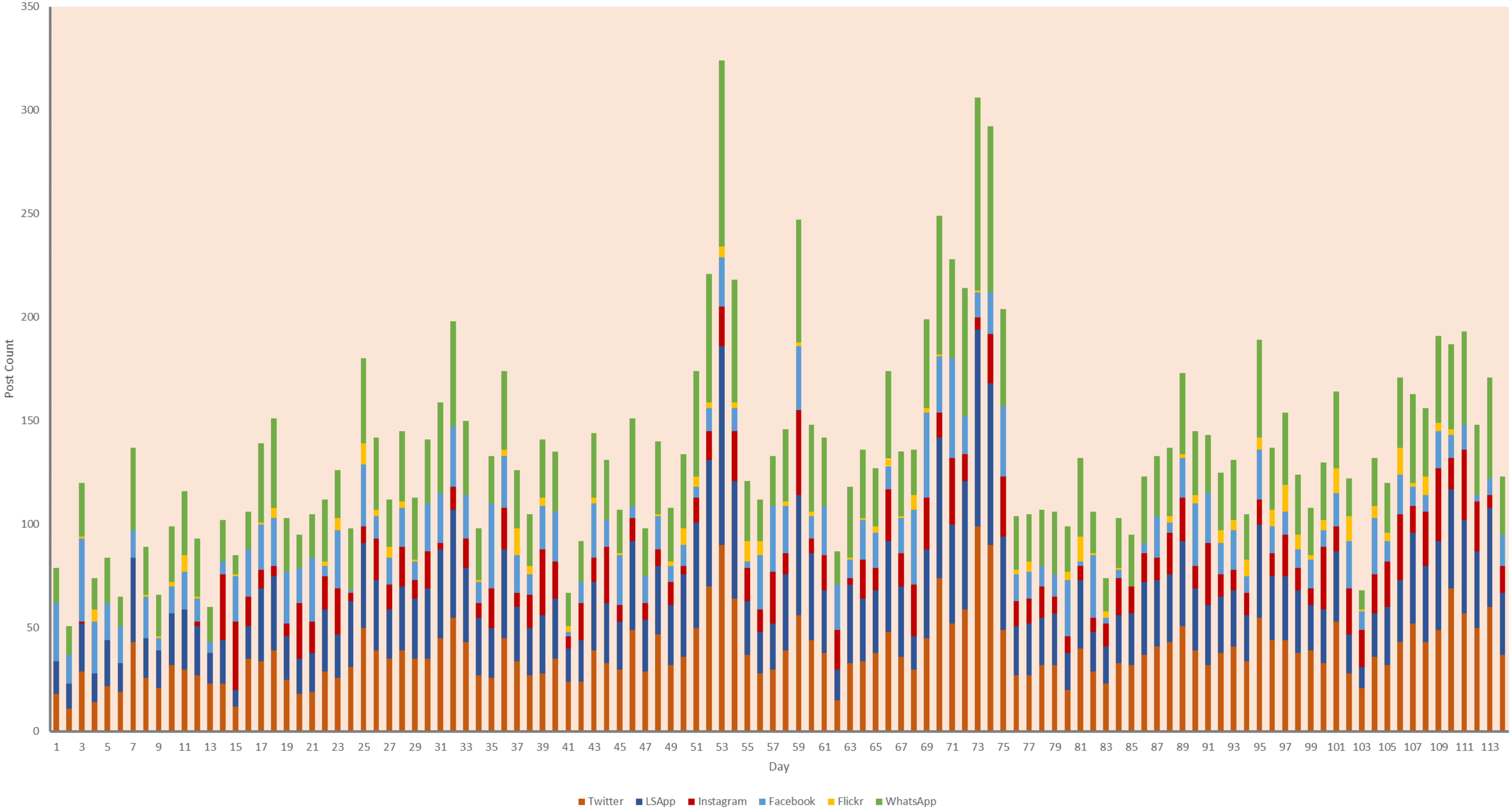

Figure 4: The number of mammal -sighting-based posts on each platform daily, over the 114-day study period which began on the 16th February 2023 (day 1) and concluded on 9th June 2023 (day 114). Observations were recorded over 150 days, but the dataset was reduced to the aforementioned range due to insufficient coverage towards the start and end of this period.

Table 5: The presence and absence of all captured species on every platform across both parks. The table is organised (and shaded) by taxonomic family in alphabetical order with species therein being sorted by the percentage of platforms they were recorded on (ascending order). The final column indicates the percentage of platforms species were recorded on (presence). The final row indicates the percentage of species each platform recorded a sighting of (coverage).

			Platform						
Family	**Common Name**	**Scientific Name**	Twitter	LSApp	Instagram	Facebook	Flickr	WhatsApp	***Presence***
Bovidae	Blesbok	*Damaliscus pygargus phillipsi*	×	×	×	✓	×	×	16.67%
Bovidae	Red Hartebeest	*Alcelaphus buselaphus caama*	×	✓	×	×	×	×	16.67%
Bovidae	Eland	*Taurotragus oryx*	×	✓	✓	×	×	×	33.33%
Bovidae	Reedbuck	*Redunca arundinum*	×	✓	✓	×	×	×	33.33%
Bovidae	Sable	*Hippotragus niger*	✓	✓	×	×	×	×	33.33%
Bovidae	Steenbok	*Raphicerus campestris*	×	✓	×	✓	×	×	33.33%
Bovidae	Duiker	*Sylvicapra grimmia*	×	✓	×	✓	✓	×	50.00%
Bovidae	Springbok	*Antidorcas marsupialis*	×	×	✓	✓	✓	×	50.00%
Bovidae	Bushbuck	*Tragelaphus scriptus*	✓	×	✓	✓	✓	×	66.67%
Bovidae	Tsessebe	*Damaliscus lunatus*	✓	✓	×	✓	✓	×	66.67%
Bovidae	Greater Kudu	*Tragelaphus strepsiceros*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Impala	*Aepyceros melampus*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Klipspringer	*Oreotragus oreotragus*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Nyala	*Tragelaphus angasii*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Waterbuck	*Kobus ellipsiprymnus*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Wildebeest	*Connochaetes taurinus*	✓	✓	✓	✓	✓	×	83.33%
Bovidae	Buffalo	*Syncerus caffer*	✓	✓	✓	✓	✓	✓	100.00%
Canidae	Side-Striped Jackal	*Canis adustus*	×	✓	×	×	×	×	16.67%
Canidae	Black Backed Jackal	*Canis mesomelas*	✓	✓	✓	✓	✓	×	83.33%
Canidae	African Wild Dog	*Lycaon pictus*	✓	✓	✓	✓	✓	✓	100.00%
Cercopithecidae	Baboon	*Papio ursinus*	✓	✓	✓	✓	✓	×	83.33%
Cercopithecidae	Vervet Monkey	*Chlorocebus pygerythrus*	✓	✓	✓	✓	✓	×	83.33%
Elephantidae	Elephant	*Loxodonta africana*	✓	✓	✓	✓	✓	×	83.33%
Equidae	Plains Zebra	*Equus quagga*	✓	✓	✓	✓	✓	×	83.33%
Felidae	Caracal	*Caracal caracal*	✓	✓	×	✓	×	×	50.00%
Felidae	Serval	*Leptailurus serval*	✓	✓	×	✓	×	×	50.00%
Felidae	African Wildcat	*Felis lybica*	✓	✓	×	✓	×	×	50.00%
Felidae	Cheetah	*Acinonyx jubatus*	✓	✓	✓	✓	✓	×	83.33%
Felidae	Leopard	*Panthera pardus*	✓	✓	✓	✓	✓	✓	100.00%
Felidae	Lion	*Panthera leo*	✓	✓	✓	✓	✓	✓	100.00%
Galagidae	Bushbaby	*Galago moholi*	✓	✓	×	×	×	×	33.33%
Giraffidae	Giraffe	*Giraffa giraffa*	✓	✓	✓	✓	✓	×	83.33%
Herpestidae	Banded Mongoose	*Mungos mungo*	×	✓	✓	×	✓	×	50.00%
Herpestidae	Dwarf Mongoose	*Helogale parvula*	×	✓	✓	×	✓	×	50.00%
Herpestidae	Black Tailed Mongoose	*Canis mesomelas*	×	✓	✓	×	✓	×	50.00%
Hippopotamidae	Hippopotami	*Hippopotamus amphibius*	✓	✓	✓	✓	✓	×	83.33%
Hyaenidae	Brown Hyena	*Hyaena brunnea*	×	✓	✓	✓	×	×	50.00%
Hyaenidae	Spotted Hyena	*Crocuta crocuta*	✓	✓	✓	✓	✓	✓	100.00%
Hystricidae	Porcupine	*Hystrix africaeaustralis*	×	✓	×	×	×	×	16.67%
Leporidae	Scrub Hare	*Lepus saxatilis*	×	✓	×	×	×	×	16.67%
Macroscelididae	Elephant Shrew	*Macroscelididae sp.*	×	✓	×	×	×	×	16.67%
Manidae	Pangolin	*Smutsia temminckii*	×	✓	×	×	×	×	16.67%
Molossidae	Little Free-Tailed Bat	*Chaerephon pumilus*	×	×	×	×	✓	×	16.67%
Mustelidae	Honey Badger	*Mellivora capensis*	×	✓	✓	×	✓	×	50.00%
Pedetidae	Springhare	*Pedetes capensis*	×	×	×	×	✓	×	16.67%
Procaviidae	Dassie	*Procavia capensis*	×	✓	×	×	×	×	16.67%
Pteropodidae	Wahlberg's Epauletted Fruit Bat	*Epomophorus wahlbergi*	×	✓	×	×	×	×	16.67%
Rhinocerotidae	Black Rhinoceros	*Diceros bicornis*	✓	×	✓	✓	✓	×	66.67%
Rhinocerotidae	White Rhinoceros	*Ceratotherium simum simum*	✓	×	✓	✓	✓	×	66.67%
Sciuridae	Smith's Bush Squirrel	*Paraxerus cepapi*	×	✓	✓	×	✓	×	50.00%
Sciuridae	Tree Squirrel	*Paraxerus cepapi*	✓	×	✓	✓	✓	×	66.67%
Soricidae	Forest Shrew	*Myosorex varius*	×	×	×	×	✓	×	16.67%
Suidae	Warthog	*Phacochoerus africanus*	✓	✓	✓	✓	✓	×	83.33%
Viverridae	Civet	*Civettictis civetta*	×	✓	✓	×	×	×	33.33%
Viverridae	Genet	*Genetta tigrina*	×	×	✓	✓	✓	×	50.00%
		Coverage	54.55%	81.82%	63.64%	61.82%	65.45%	9.09%	

What became evident when observing these numbers is that the number of posts about a species was not proportional to the number of sightings. Leopards, for example, were subject to 2.72 posts for every sighting made (a repost ratio of 1:2.72). The three big cats (leopard, cheetah, lion) all experienced the highest repost rates, alongside elephant, as presented in table 6. With the exception of buffalo and elephant, herbivorous species received significantly fewer reposts than carnivores. Elephant and buffalo are among a group known as the Big Five; animals infamous for the danger they pose to hunters, while antelope, giraffe, and zebra are not. For non-Big Five herbivorous species, there was a substantial difference between repost ratios and those of carnivores, buffalo, and elephant, with zebra, they 8th most reposted species, experiencing 30% less reposts than spotted hyena, the least reposted carnivore.

Table 6: The repost ratios (the number of posts generated per individual and unique sighting) of the ten most prevalent species, ordered alphabetically.

Species (Group)	Repost Ratio	Number of Posts	Number of Sightings
Antelope	1.32	1035	785
Buffalo	2.04	926	453
Cheetah	2.29	917	425
Elephant	2.29	2,449	1,071
Giraffe	1.35	786	583
Leopard	2.72	2,406	884
Lion	2.37	4,827	2,040
Spotted Hyena	2.01	1087	540
Wild Dog	2.33	938	402
Zebra	1.47	506	344

It was evident that the rate at which a species experienced reposts was not proportional to the number of sightings generated for it, nor did reposts serve to increase the prevalence of any species through posts above its representation through sightings. Elephant, for example, was the second most sighted and posted species, yet the repost ratio for elephants was the fifth highest (1:2.29), tied with cheetah which was the eight most sighted and posted species. Wild dog, the sixth most sighted and posted species, had a higher repost ratio than elephant and cheetah at 1:2.33; indicating that more frequently sighted species are not subject to more reposts, and that species popularity among posts is not a result of said reposts.

4.1.2 Species Community Composition on Platforms

There was a statistically significant association between species community composition and platform (X^2=1,322, *d.f.* =45, p <0.001). Figure 5 represents the counts of sightings and posts for each species by platform. Lion was the most sighted and posted species on Twitter, LSApp, Facebook, and WhatsApp, and elephant was the most sighted and posted species on Instagram and antelope was the most sighted and posted taxa on Flickr. The second and third most sighted/posted taxa were typically either elephant or leopard, with the exception of Facebook where elephant was second and antelope was third, and Flickr where elephant was second and giraffe was tied third with lion.

A Bray-Curtis measure of community dissimilarity with all species across all parks pooled generated a score of 66.5, indicating considerable variation in community composition across platforms, regardless of park. Table 7 presents a matrix of Bray-Curtis scores for comparisons of pairs of platforms, where the average dissimilarity among pairs was 74.37.

Table 7: Bray-Curtis' measure for community dissimilarity scores from comparisons of platforms as pairs of environments. Averages indicate the mean dissimilarity of a platform to all others.

	Twitter	LSApp	Instagram	Facebook	Flickr	WhatsApp	Average
Twitter		70.79	60.07	60.79	73.13	80.24	69.00
LSApp	70.79		61.32	61.32	85.88	90.80	74.02
Instagram	60.07	61.32		52.94	73.27	86.35	66.79
Facebook	60.79	61.32	52.94		74.65	86.46	67.23
Flickr	93.13	85.88	73.27	74.65		87.50	82.89
WhatsApp	80.24	90.80	86.35	86.46	87.50		86.27
Average	69.00	74.02	66.79	67.23	82.89	86.27	74.37

Flickr and WhatsApp were the most dissimilar platforms, with average Bray-Curtis scores of 82.89 and 86.27 respectively. Both platforms were most dissimilar to LSApp and one another, but the high dissimilarity scores indicate that the compositional divergences between them and LSApp are different to those between each other. Conversely, Instagram and Facebook were the least dissimilar platforms compared to all platforms, and also with each other; possessing a dissimilarity score of 52.94, indicating similar community compositions.

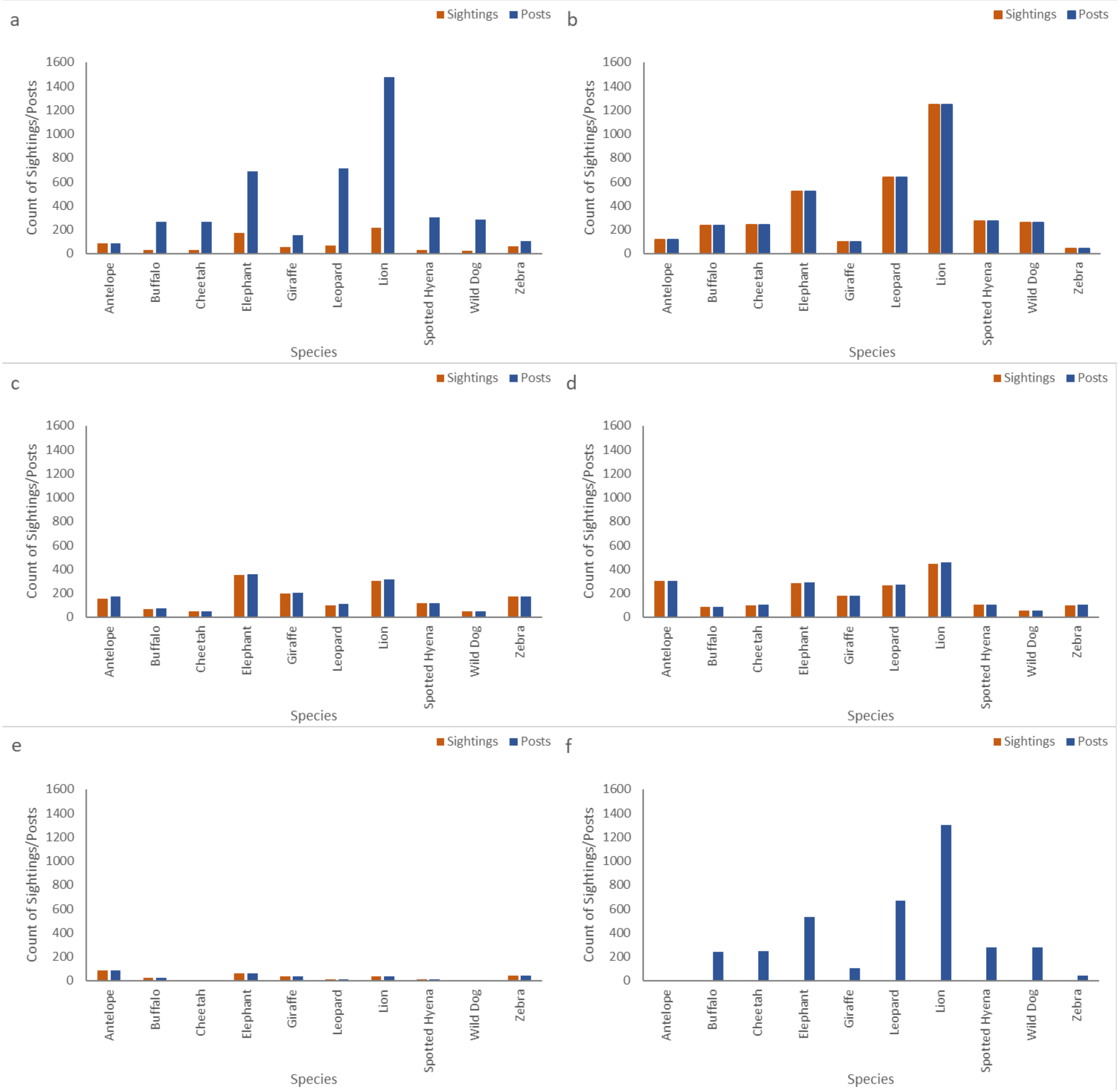

Figure 5: Sighting and post counts for species across platforms; KNP and PNP pooled. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

A pair-wise one-way PERMANOVA applied to the data determined that the difference in representations of taxa between platforms was statistically significant (Σ^2=118, *d.f.* =37, *p* <0.001). This pairwise contract of *p* <0.001 indicates that the differences in representation of a single taxa across multiple platforms was varied and statistically significant. A Chi-square test for statistical association, on the other hand, determined that the representations of all taxa across individual platforms were statistically significant (X^2=1,322, *d.f.* =45, *p* <0.001) and that there is an association between a platform and its community composition.

Within both the pooled Bray-Curtis measure and all instances of platform pair comparison, lion was the greatest source of community dissimilarity. Lion contributed 24.46% to the pooled dissimilarity score while elephants, the second most dissimilar, contributed 14.60%. Cheetah was the least dissimilarly represented species, contributing just 5.45% to the pooled Bray-Curtis score.

This indicated a correlation between the number of sightings a species receives and its contribution to community dissimilarity. This was likely due to the overrepresentation of these species by specific platforms, relative to their sighting counts of other species. Lion on LSApp, for example, were subject to an average of 10.90 daily sightings, whereas all other platforms sighted lions just 2.07 times per day. As can be seen in figure 5, lions were sighted many more times than any other species on LSApp, but this difference between most sighted and second most sighted species is substantially less on other platforms.

4.1.3 Species Accessibility across Platforms

The effort required to observe the first instance of a species varied by platform and by species. Facebook stood out as the platform where sightings most quickly recorded all top species, doing so in the first day of data capture. Twitter and LSApp both took until day 6 to first sight all species, Instagram until day 13, and Flickr until day 53 (though only until day 21 to report all species other than cheetah). WhatsApp failed to detect all of the ten most prevalent species as independent sightings. Figure 6 presents these accrual rates by sightings, and figure 7 for Twitter and WhatsApp by posts.

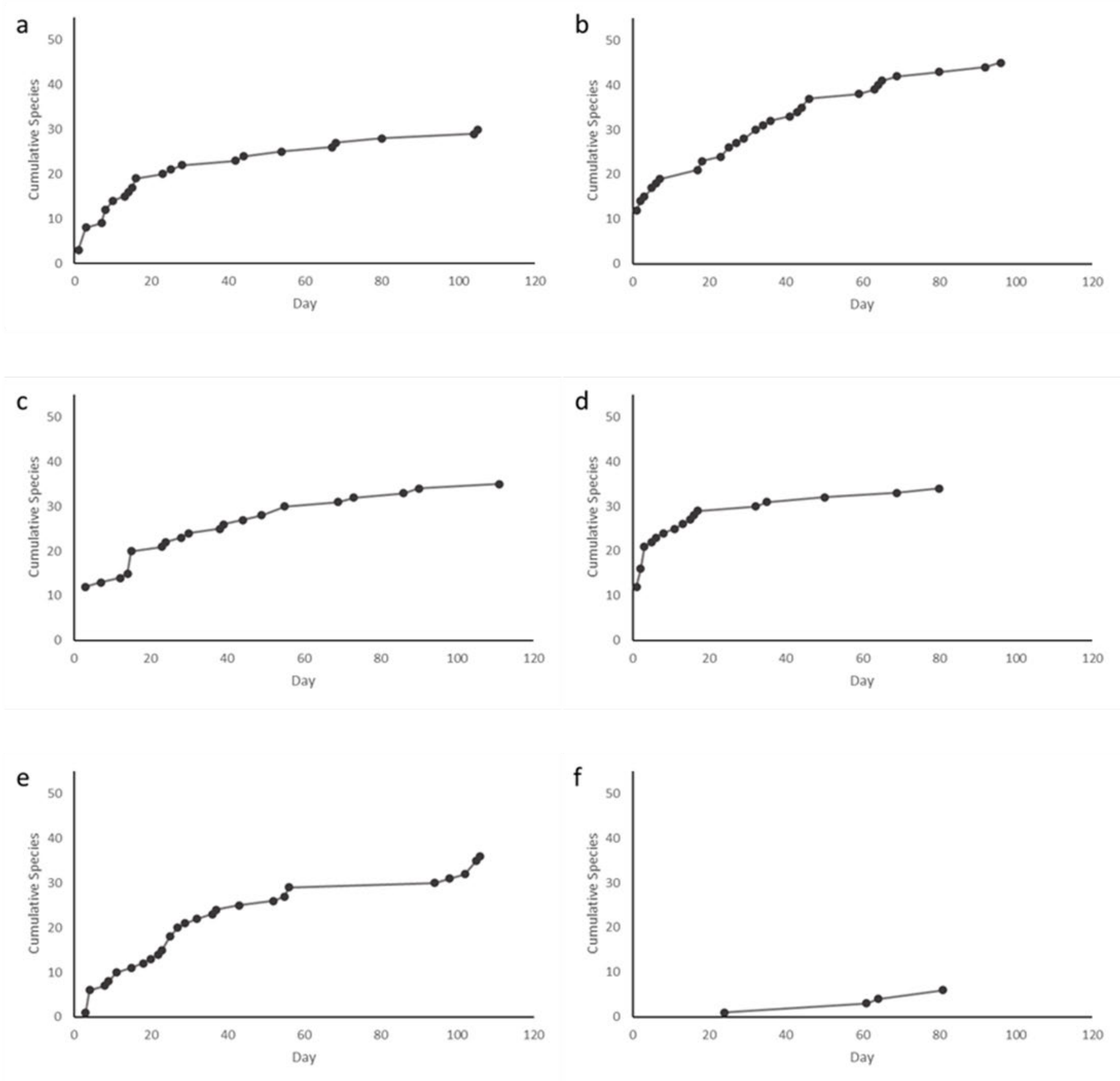

Figure 6: Species Accumulation Curves by platform, based on the day on which the first sighting of a previously unrecorded species was made. X-axis maximum 114 is the range of the study, y-axis maximum 55 is the total number of species observed. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

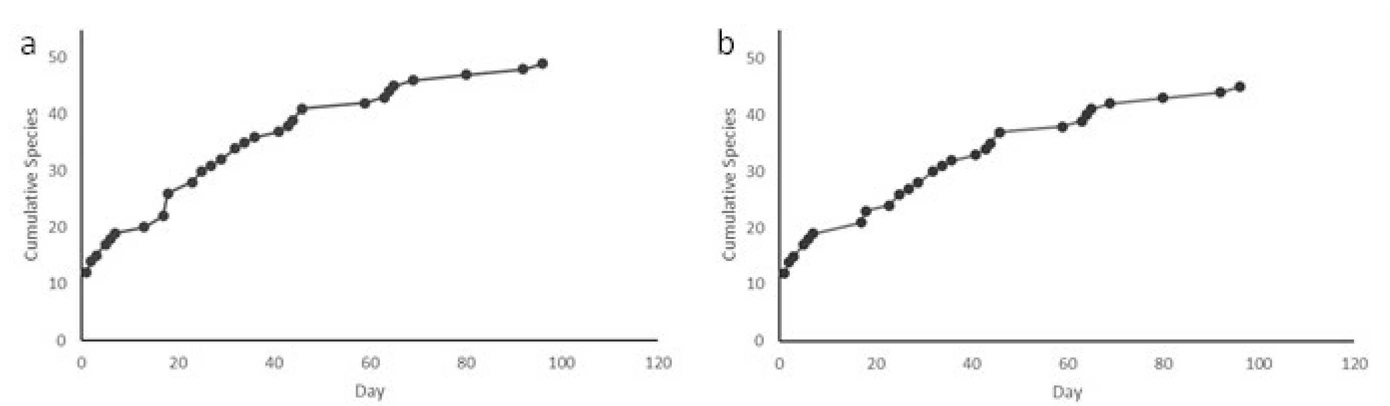

Figure 7: Species Accumulation Curves by platform, based on the day on which the first post of a previously unrecorded species was made. X-axis maximum 114 is the range of the study, y-axis maximum 55 is the total number of species observed. (a) Twitter, (b) WhatsApp

There was significant variation in the accrual rates of previously unsighted species across platforms. Facebook's SAC (fig 6d) represents the most typical expression of the graph where a logarithmic curve is expressed due to a rapid rate of accumulation at the beginning of the observation period and reduced opportunity for new species to be sighted further on slowing the rate of accumulation. Facebook also recorded the most species by day 20; accruing 29 unique species while by the same time LSApp (fig 6b) had only sighted 21. LSApp would proceed to accumulate more species, however, would do so in an irregular manner. Jumps in rates of accrual at days 23-25, 29-34, and 44-46 represent the sighting of Smith's bush squirrel, honey badger, Wahlberg's epauletted fruit bat, banded mongoose, dwarven mongoose, scrub hare, brown hyena, steenbok, duiker, and waterbuck. Notably, with the exception of waterbuck, all of these species would be hard to sight due to (1) size; (2) abundance; (3) time of activity; (4) accessibility of preferred habitat by OCs, but many were sighted on other platforms sooner. LSApp and more so Flickr indicate a potential to detect more species than observed here, with new species being detected late into the 114-day observation period.

Figure 7 illustrates the meaningful impact of LSApp reposts on Twitter and WhatsApp. WhatsApp's SAC replicates LSApp's when posts are considered, due to no new species being detected among the few WhatsApp sightings and all mutually detected species being reported by LSApp sooner. Twitter on the other hand presents slight variation where some mutual species were detected sooner on Twitter and some entirely new species were covered which LSApp did not report. This is most notable between day 15 and day 40, where all four of the species Twitter sightings reported that LSApp did not were first sighted.

Species accumulation was not only determined by platform, but also dependent on the recorded taxa. Figure 8 illustrates the amount of effort needed (number of days taken) to detect a species, on average, across all platforms. Species with higher sighting counts were typically sighted sooner, with elephant and lion being the species for which sightings were most quickly made. Species abundance also possibly played a role with cheetah and wild dog taking the longest of any most prevalent species to be detected, and those species having the smallest populations in the parks.

4.1.4 Parks

Despite the composition of posts made by each account type across each park being presented as similar in figure 9, there was a statistically significant association between account type and park (X^2=10, *d.f.* =2, p =0.005986942). This would indicate a level of diversity between the most active account types within each park. Figure 10 demonstrates the usage of different platforms within each park, which appears to be much more significantly diverse (X^2=325, *d.f.* =5, p <0.001). It therefore seems that the apparent similarity of account type composition within the parks is not a direct consequence of account type, but rather a consequence of more varied platform usage, and the account type variation upon said platforms, within parks, illustrated in figure 11.

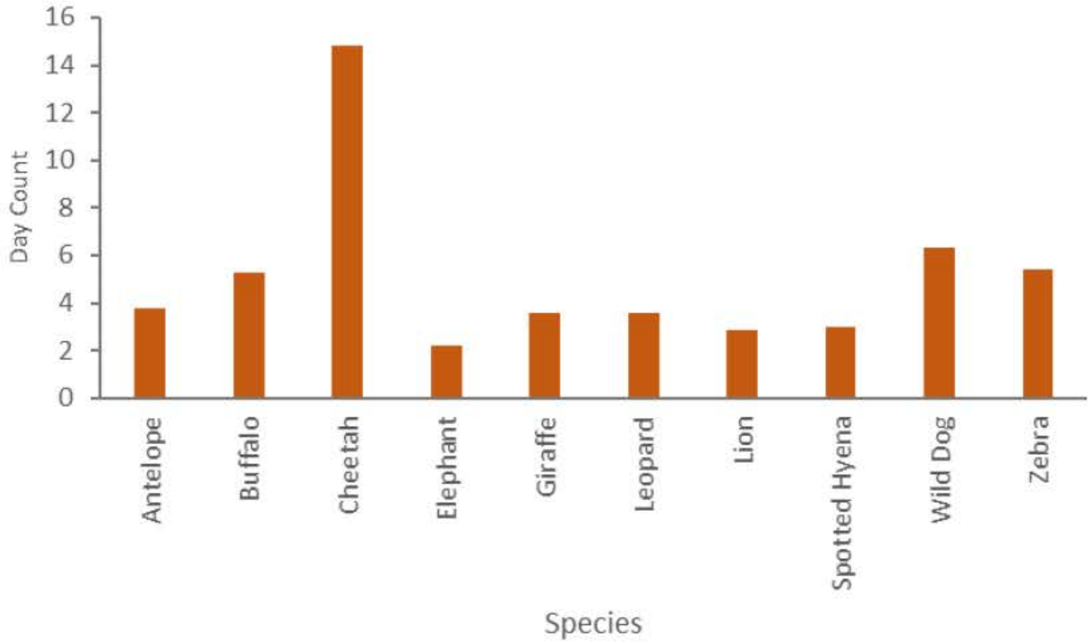

Figure 8: The average number of days before the first sighting of a most prevalent taxa was made across all platforms apart from WhatsApp, which was excluded as it did not report independent (un-reposted or multiposted) sightings all of the most prevalent taxa.

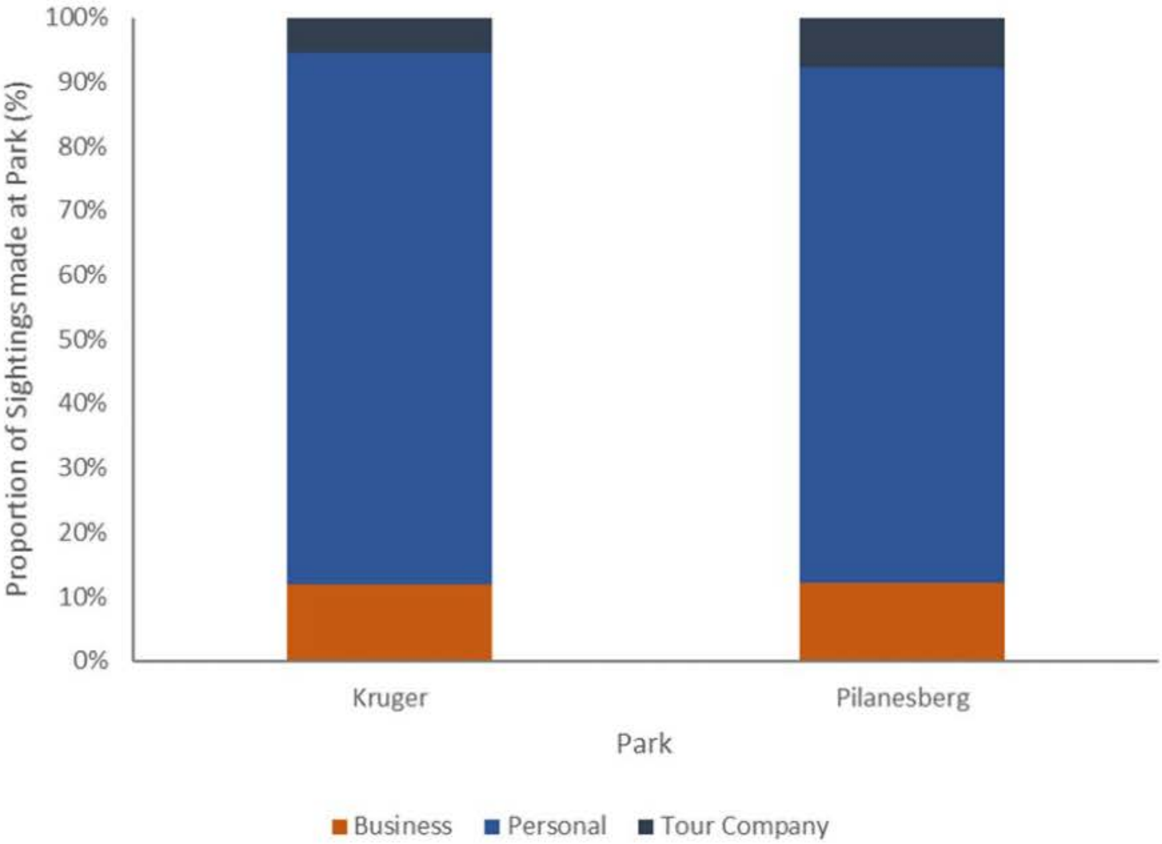

Figure 9: The distribution of account types across both parks based on their contribution to instances of unique sightings.

There was a significant association between park and platform usage (X^2=325, *d.f.* =5, p <0.001), outlined in figure 10. LSApp was most used in both parks, while WhatsApp reported the fewest sightings. Facebook was more popular among PNP visitors and Instagram less so, while Twitter and Flickr experienced similar usage in both parks.

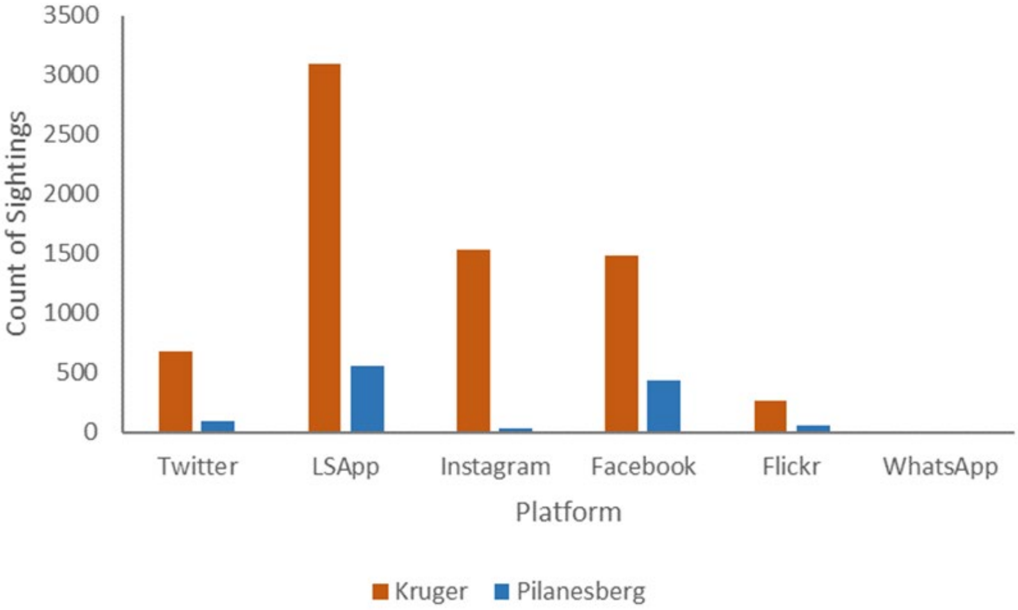

Figure10: The count of sightings from each park uploaded to platforms.

Figure 11 illustrates the trends between platform and account type across parks, showing a statistically significant association between platform and account type overall (X^2=2396, *d.f.* =10, p <0.001). Twitter users in KNP were mostly business accounts, with little business presence on LSApp, Instagram, or Facebook. Flickr sightings for both parks were exclusively from personal accounts. Tour companies were prevalent on LSApp, accounting for 12% of KNP sightings and 15% of PNP sightings. Twitter, Instagram, and Facebook users presented significant differences between parks. Twitter users in PNP were mostly personal accounts, while businesses were more active on Instagram and Facebook than in KNP. The correlation between account type and park was the result of the physical locations and the complex interactions between account types and platform, and subsequently between platform usage and parks, creating compositions that at first appeared convergent.

Spotted hyena was the only species among the ten most prevalent not sighted in PNP due to their known absence. In KNP, lions were the most frequently sighted species with 2,070 sightings, followed by elephant with 1,090 sightings. In PNP, this order was reversed, with elephants being the most sighted species (331) and lions coming second (235). Leopard ranked third in both parks, but the difference in sightings between leopard and lions was much smaller in PNP (217 vs. 235) than in KNP (893 vs. 2070), relative to their frequency.

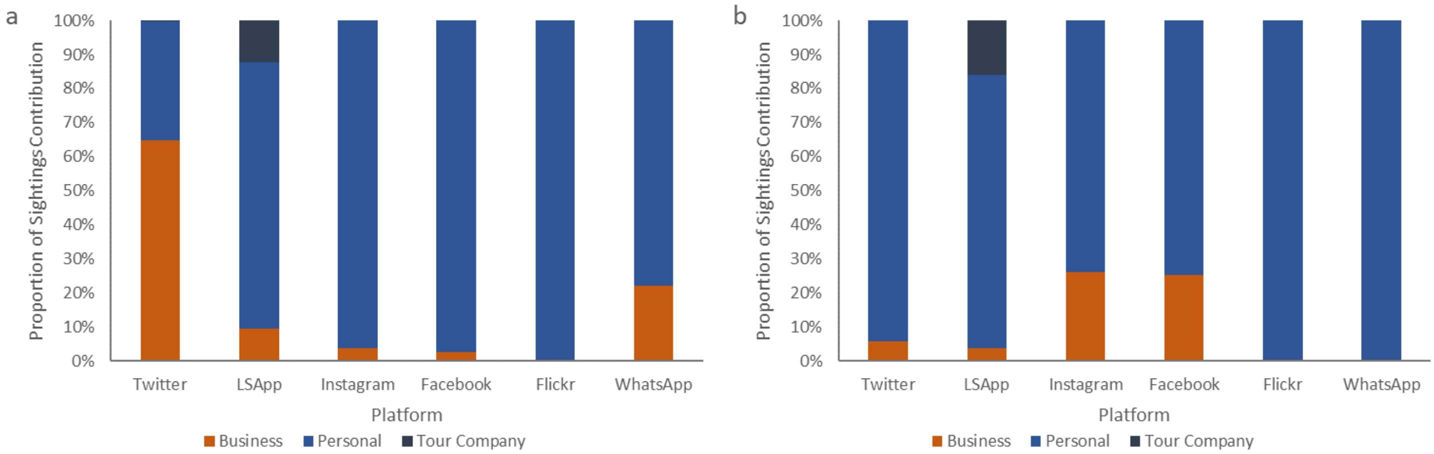

Figure 11: The contribution to total sightings counts of each account type across platforms, divided by park. (a) Kruger, (b) Pilanesberg.

The prevalence of all other species varied between parks, also, except for cheetah, which accounted for approximately 5% of sightings in both parks. There was a statistically significant association (X^2=367, *d.f.* =9, *p* <0.001) between species and park, suggesting differences in the relative species abundance between parks and/or that variation in platform usage frequencies within parks had influenced the representation of species.

4.1.5 Account Types

Of the 8,222 posts identified to have been made by business accounts, only 986 sightings were confirmed to have been originally uploaded by the account type. This indicates that the majority of business posts are reposts, largely due to the Latest Sightings team sharing content from LSApp to Twitter and WhatsApp. In contrast, personal accounts contributed 6,874 posts from 6,846 sightings, meaning that only 28 (0.4%) of personal account posts were not oc. However, personal accounts were responsible for 82% of all multiposts.

Tour companies generated 471 posts, mostly consisting of unique sightings. Tour companies were only distinguishable from regular companies on LSApp, where their usernames were italicized. This meant that their sightings also contributed to the repost numbers on Twitter and WhatsApp. Rarely, the same sighting was multiposted by multiple tour companies on LSApp.

In addition to park and platform, account type also showed a statistically significant association with species (X^2=284, *d.f.* =18, *p* <0.001). Business accounts generated the most posts for most taxa, except for antelope, giraffe, and zebra where personal accounts contributed the most posts. However, when it came to sightings data, personal accounts generated the most sightings for all taxa. Tour companies generated the fewest sightings and posts for all taxa, with leopard and lion being the most popular among tour companies.

Only 3 sightings were subject to multiposting by tour companies, and all of them were lion sightings. Among personal accounts, lion, leopard, and elephant were the subject of the most reposts. Antelope, buffalo, and wild dog did not receive a single repost from personal accounts. Business account posts were often duplicates of sightings from other business and personal accounts. Just 11.7% of business posts were original sightings from the group. Business accounts reposted and shared a higher concentration of business posts than of any other account type. 90% of all cheetah and lion business sightings were the result of reposted content. Buffalo were the only species reposted more frequently by business accounts. Personal and tour company account habits both differed slightly, with leopard, lion, and cheetah being the most reposted species among them. The least reposted species by business accounts was zebra, where 60% of business posts were original business account zebra sightings.

Furthermore, there was a strong, statistically significant association between account type and content type (X^2=2893, *d.f.* =6, p <0.001). Personal accounts were less likely to create text-based posts but more likely to share image, long-form video, and video posts. Tour companies were more likely to upload text-based sightings, while business accounts were the least likely to produce long-form video content. Personal accounts were dominated by image content, accounting for 71% of all sightings. Business accounts had 55% of sightings posted as images, and 39.5% as text. Similarly, 95.2% of sightings from tour companies were split between image and text (45.64% and 49.68% respectively).

4.1.6 Content Type

As presented in figure 12, image was by far the most prevalent means of recording an animal for every taxa. On average, 75.02% of sightings were recorded as images, with just 13.20% of sightings being recorded as text-based posts, 10.93% as video, and 0.85% as long-form videos. This, however, varies with taxa, presenting a statistically significant association (X^2=341, *d.f.* =27, p <0.001). A decrease in image sightings seemed to correlate with sighting frequency. Highly sighted species like lion, leopard, and elephant had more diversity in the content types they were represented by than zebra, antelope, and giraffe.

Content type also corresponded to platform (X^2=2617, *d.f.* =15, p <0.001). For Twitter, Instagram, and Facebook, image content was the most prevalent with video content the second. Flickr also covered the majority of its sightings through images, with minimal secondary contribution from video. LSApp and WhatsApp were the only platforms where text content was the second most prevalent content type. Long-form videos were significantly more prevalent on Facebook than anywhere else, as shown in figure 13.

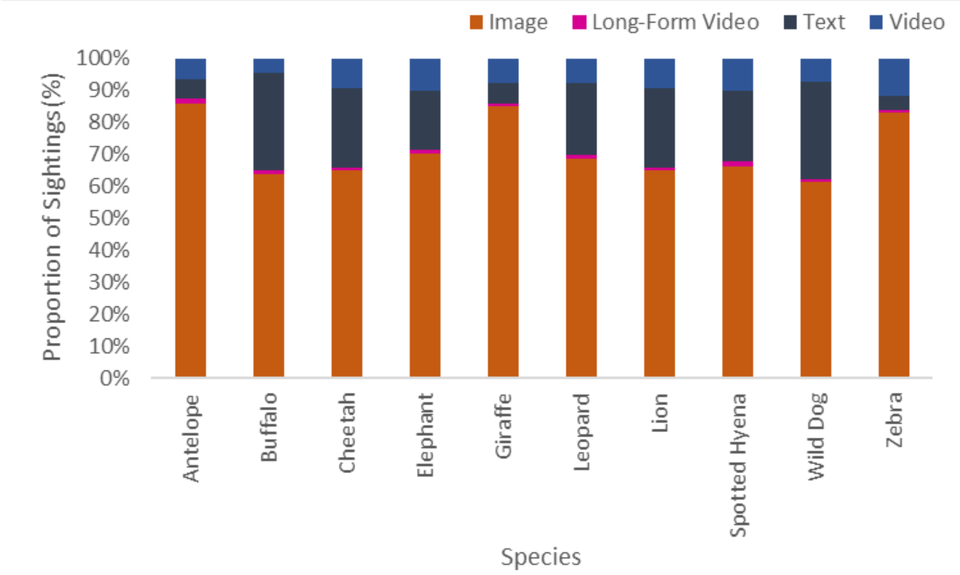

Figure 12: The proportion of sightings of each taxa represented as each content type. KNP and PNP pooled.

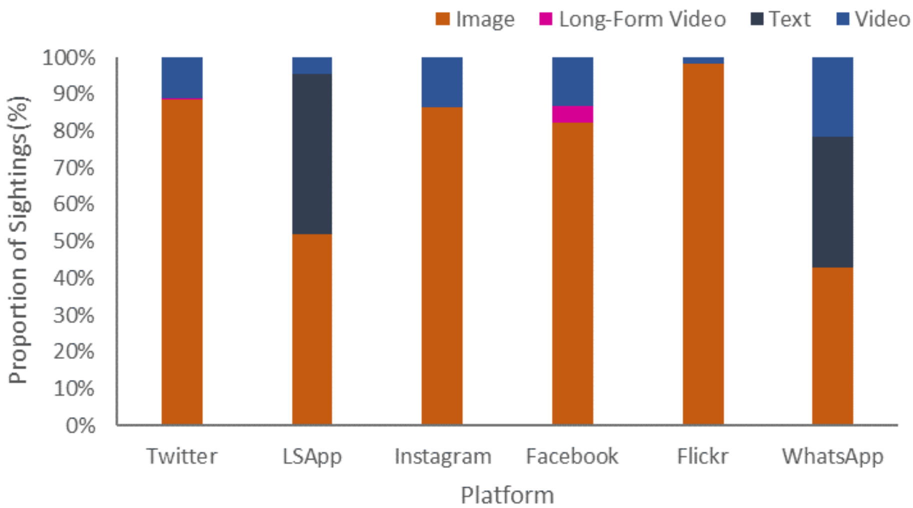

Figure 13: The proportions of sightings made to each platform represented as each content type. KNP and PNP pooled.

Figure 14 presents the variations in content type across platform relative to lion sightings. There was a statistically significant association between the representation of lion sightings across platforms (X^2=622, *d.f.* = 15, *p* <0.001). For lion sightings on Twitter, Instagram, Facebook, and Flickr, image and video posts were the most common content types, while LSApp and WhatsApp had a notable presence of text-based sightings, corresponding with the representations of all species across all platforms.

Leopard sightings were similarly distributed, though with an absence of text-based sightings on WhatsApp (see Appendix A, figure A1). There was an increased prevalence of image sightings for antelope, with video sightings being the second most common. There were no antelope sightings on WhatsApp (see Appendix A, figure A2).

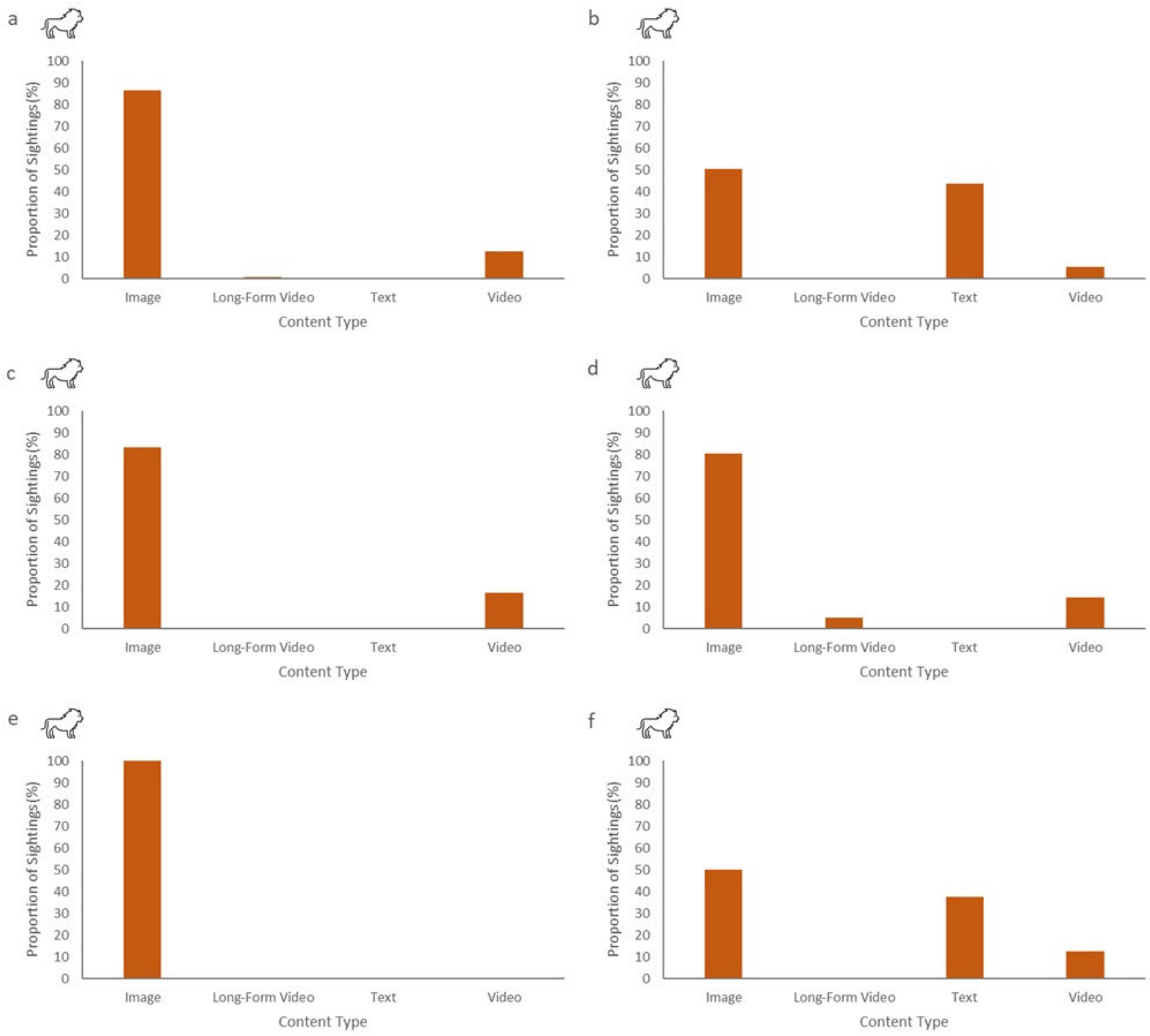

Figure 14: The distribution of lion sightings by content type across platforms. KNP and PNP pooled. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

4.1.7 Animal Behaviour and its impact on sightings and posts

Ethology influenced many aspects of how sightings were documented. Due to the differing activity budgets of taxa, tests were only applied to lion, leopard, and antelope to determine what impact their behaviour had on where, how, and by whom a sighting was reported.

There were statistically significant associations between behaviour and platform for lion, leopard, and antelope (lion, X^2=459; leopard, X^2=268; antelope, X^2=105; all *d.f.* =45, all p <0.001). Lion and leopard sightings shared similarities across platforms, with resting and locomotion behaviours being the most prevalent. Antelope sightings were more varied and presented feeding, mating, and portrait most commonly. Sightings of all taxa experienced an increase in representation as portrait content on Instagram, Facebook, and Flickr, which exhibits the priorities of the OC rather than the behaviours of the animals. Overall, representation across platforms seem to correspond with the expected behaviour budgets of these taxa. Figure 15 represents the slight variations in behaviour across platforms (see Appendix A figure A3 for leopard and A4 for antelope).

Facebook presented the most diverse recordings of behaviours, with Twitter recording the second most diversity for leopard, and Flickr doing so for antelope. Interspecies interactions were rarely reported outside of Facebook and WhatsApp was the most common source of sightings without confidently discernible behaviour. Despite representing other behaviours the same as other platforms, Instagram presented feeding behaviours for lion less than other platforms. This discrepancy is greater among leopard sightings (see figure 16a) and inversed for antelope (see figure 16b).

Reposts from LSApp notably influenced the representation of behaviour on Twitter and WhatsApp through posts, aligning them with LSApp's compositions. Otherwise, regarding posts instead of sightings presented few changes to behaviour representation on other platforms; the only other notable difference being a further reduction of lion and leopard feeding representation on Instagram due to other behaviours (particularly portrait) being more consistently multiposted.

Additionally, animal behaviour not only influenced where sightings were shared, but also what content type they were recorded as. There were statistically significant associations between behaviour and content for lion, leopard, and antelope (lion, X^2=494; leopard, X^2=556; antelope, X^2=277; all *d.f.* =24, all p <0.001). Resting and locomotive behaviours were most common for lion and leopard in image, text, and video posts, while long-form videos mostly captured locomotion or interspecies interaction events. Antelope, however, were most commonly recorded feeding through image and long-form video sightings and resting through text or video sightings. These trends are presented for lion in figure 17, and for leopard in Appendix A figure A5 and antelope in Appendix A figure A6.

There were statistically significant correlations between all most prevalent species and behaviour (X^2=1564, *d.f.* =81, p <0.001). This would, of course, be expected due to the variation in activity budget of all species, however of note is the increased prevalence of locomotion behaviours for all species recorded by this study compared to past ethological assessments. Additionally, there were clear correlations between species and prevalence of portrait sightings across all platforms (X^2=106, *d.f.* =9, p <0.001). Figure 18 presents these trends.

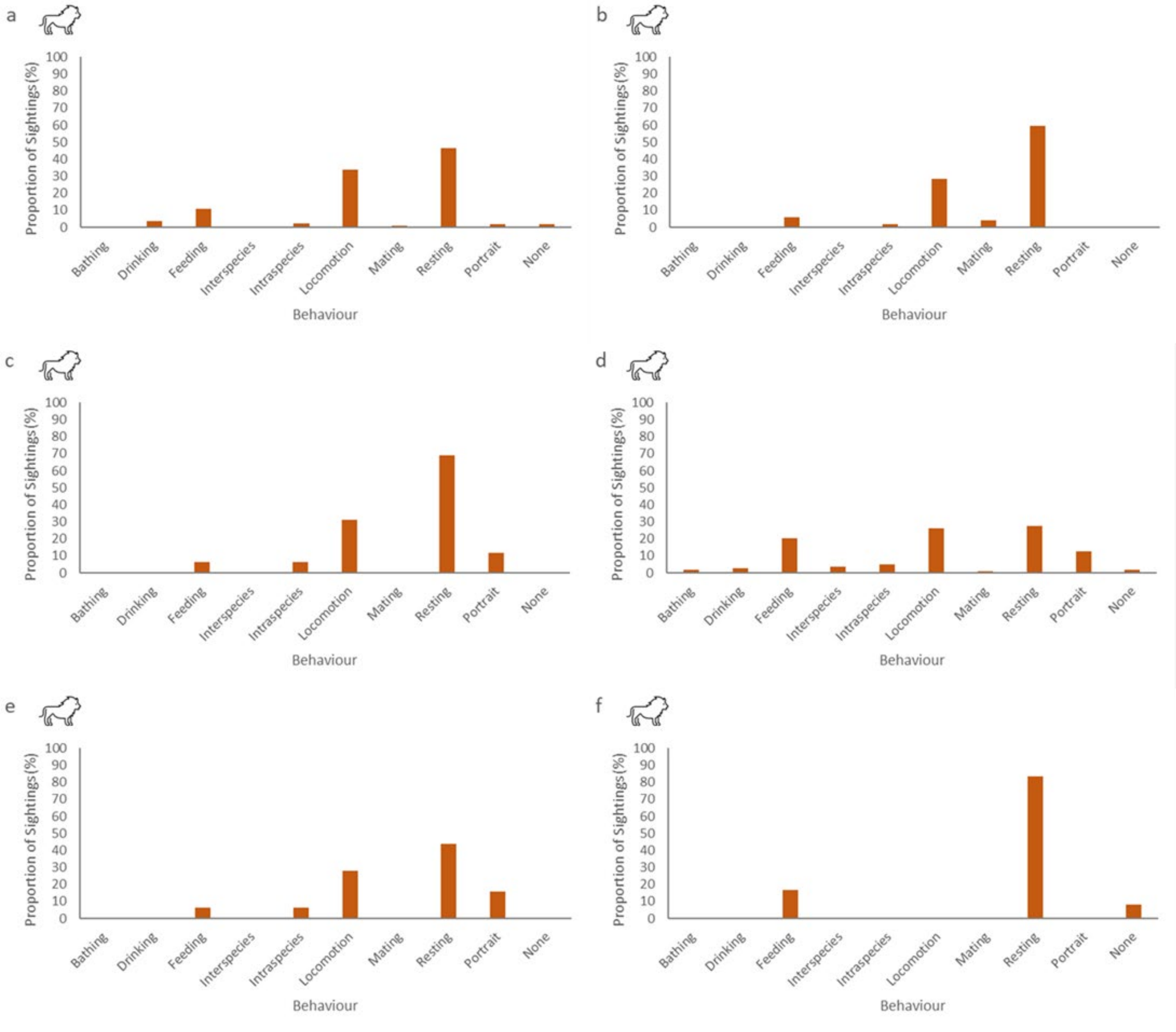

Figure 15: The distribution of lion sightings on platforms based on behaviour, represented as a proportion of total species sightings on that platform across both parks. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

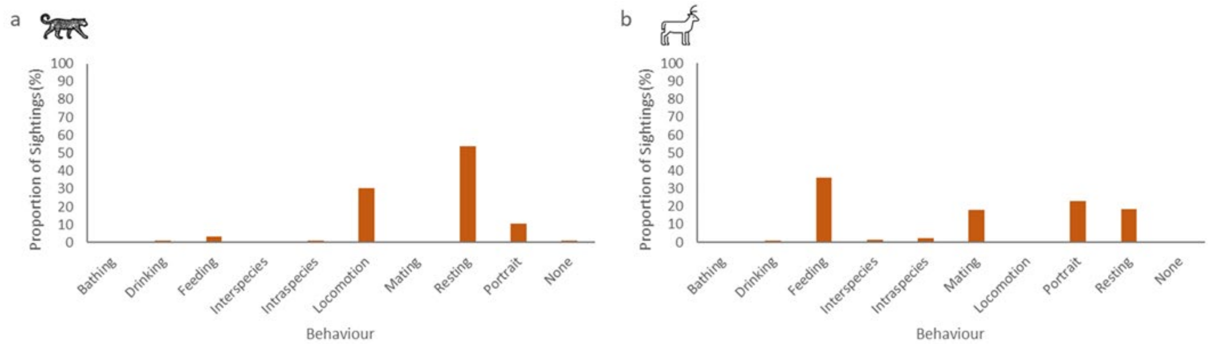

Figure 16: The distribution of sightings on Instagram based on behaviour, represented as a proportion of total taxa sightings on that platform across both parks. (a) represents leopard (which is the same graph as Appendix A figure A3c), and (b) represents antelope (which is the same figure as Appendix A figure A4c).

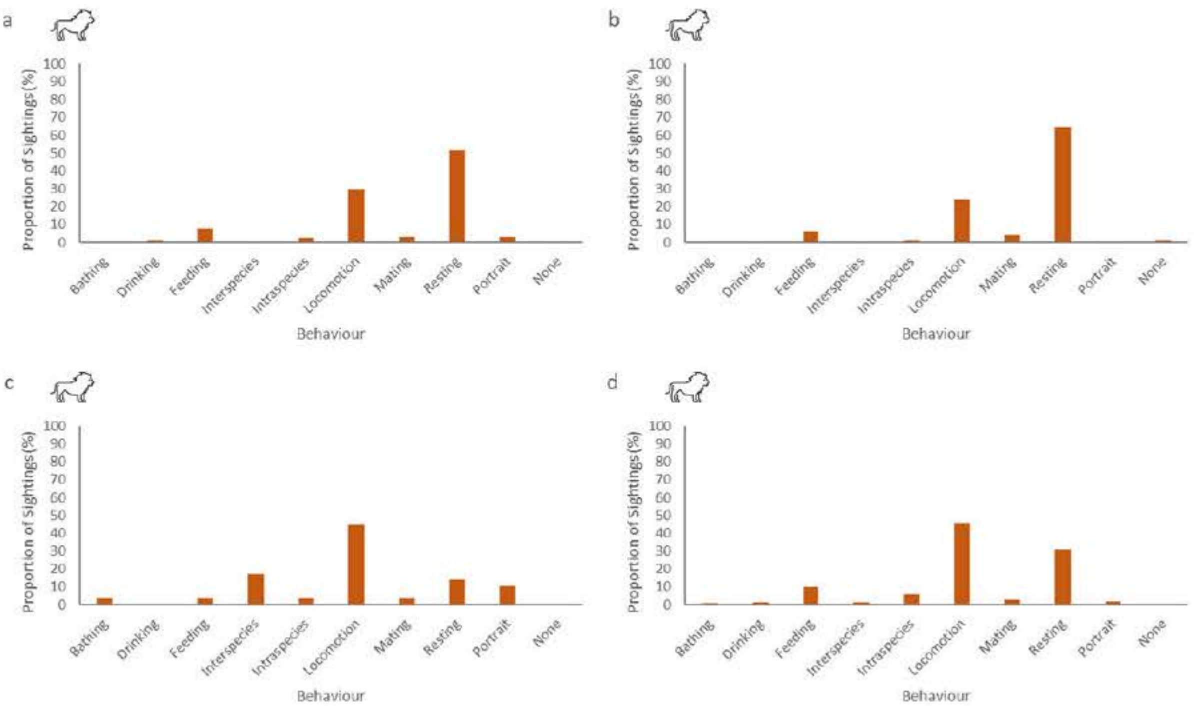

Figure 17: The distribution of lion sightings based on the expression of their behaviour through content type across all platforms and both parks. (a) image, (b) text, (c) long-form video, (d) video.

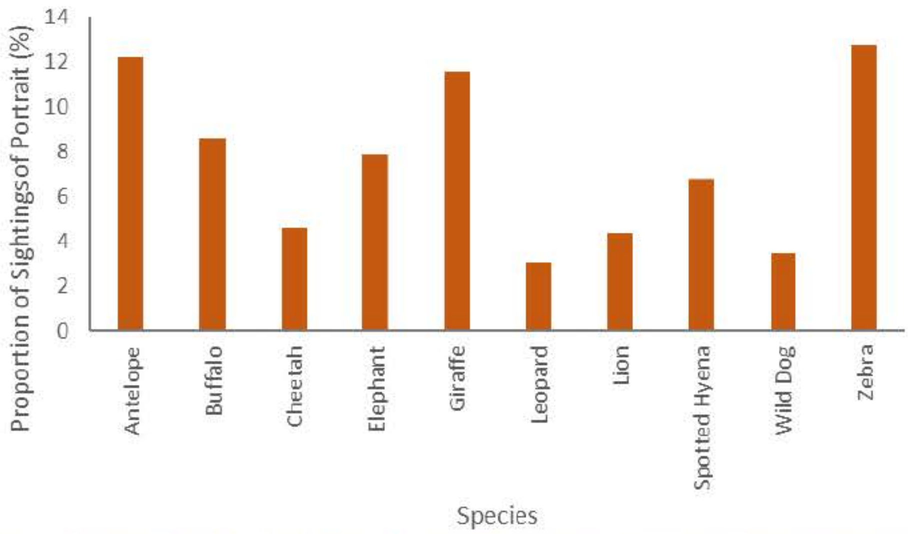

Figure 18: The proportion of all sightings of a species which were portrait sightings. Parks pooled.

4.1.8 Sighting Time

Similar to behaviour, differences between species activity budget, specifically what time of day they are active, impacts the representation of species over the course of a day. Crepuscular species, for example, are most active at dusk and dawn, and to avoid the midday heat, the majority of species in KNP and PNP would exhibit reduced intensity of activities such as feeding and locomotion around noon, where they would instead seek shade and rest.

With these concepts in mind, it was found that all most prevalent species exhibiting any behaviour were more frequently recorded in the mornings (06:30-12:00) than the afternoon and evenings (12:30-18:30). Twice as many sightings were made before noon as after. Furthermore, text-based sightings were more common than any other content type before 09:00. Locomotion, feeding, and resting behaviours were all most frequently reported at 10:00, with secondary peaks at 16:00. The most commonly reported first sighting on any day was black-backed jackal, typically before 06:00 and reported by a tour company account.

The time of day at which a sighting was reported did not appear to have an impact on its propagation as a repost.

4.1.9 Miscellaneous Observations

Four points of data were recorded which possessed trends with multiple other variables. Post lag, location, search term, and additional species all played significant roles in the accessibility of a post but did not correlate with species or behaviour; indicating that subject matter was not the driving factor of their existence or propagation.

The majority of sightings with post lag came from LSApp and Flickr where the majority of sightings with sighting day data were observed. LSApp required users to include this date, while Flickr accessed image metadata to obtain it. Of all sightings with sighting day data (48% of total sightings), almost all (99.67%) were from LSApp or Flickr. The median post lag was 0 days, with a mean of 1.85 days, and a longest post lag of 91 days. A larger proportion of posts contained post lag, raising the average to 1.92 days. This was often due to LSApp sightings from the evening being reposted the next morning, pending manual review. Some posts implied post lag in the text description through statements such as “back from holiday” or “from my time in KNP/PNP,” but could not be quantified. Efforts were made to exclude a lag of more than 6 months between sighting and upload, as these would be better suited for historical assessment. Few instances of such sightings were found.

Location data will be discussed in greater detail in section 4.4, but of note is that the majority of location data provided little tangible applicability. Location data were mandatory on LSApp, though censored for at-risk species like pangolin. Otherwise, location data were rarely provided in greater detail than a park level. When location data were provided, it was in the form of directions and landmarks, as opposed to coordinates. LSApp’s ting map was the exception to this and was the only location data deemed to be credible and specific enough to be applied; hence its integration into section 4.4. Otherwise, location data were primarily reported on Flickr by a single user whose sightings were particularly prominent on the platform and who followed LSApp's text location format.

The use of different search terms also presented correlations with content type on Twitter, Instagram, and Facebook. Search terms using #hashtags and initialisms such as (#)KNP or (#)PNP returned image and video posts, whereas phrases such as "KNP/PNP sighting" or "Kruger/Pilanesberg wildlife" returned all content types, but also returned the majority of text posts collected, due to the keywords search features of Twitter and Facebook. Search terms on Flickr acted differently than other platforms, due to the level of detail available in the default search features and the less narrative text-descriptions attached to sightings on the platform. "Kruger/Pilanesberg National Park" was a sufficient search term to yield highly concentrated, relevant information on the platform accompanied by filters on the uploaded/taken on dates, but #hashtags and initialisms were ineffective here. LSApp and WhatsApp groups did not support or need search terms due to their commitment to relevant data.

Finally, the presence of additional species in (the background of) a sighting/post was recorded. Just 1.72% of posts contained a secondary species; 0.15% of total posts a tertiary species, and 0.03% additional species beyond those. 67.53% of secondary species were recorded among image posts, and 52% of tertiary species. 60% of all additional species beyond those were reported in text posts. Antelope (predominantly impala, wildebeest, or greater kudu) was the most commonly sighted secondary species, accounting for 76.2% of secondary species. No other top species was recorded as a tertiary species, and no top species was recorded as a fourth or fifth species in a post. Tertiary species were most commonly warthog (80%) or baboon (19%). Fourth species were almost exclusively warthog, and there was one instance of baboon being identified in an image post (sat in a tree in the background of an image post, identified here but not mentioned in the text description).

4.2 Species Abundance Estimates

The abundance estimates obtained using the Lincoln-Peterson index showed an overrepresentation of all the most prevalent species in KNP. In table 8, it is evident that not only were species overrepresented in terms of count, but that they were disproportionately represented based on their contribution to the park's total population. Figure 19 illustrates these trends for nine of the most prevalent species, with the species group antelope (which accounted for 65.53% of the social media estimate and 70.37% of the census estimate) removed to increase legibility. Census data from

2012 was sourced from Roocroft (2021), with carnivore counts being verified by Wray's (2023) more recent study.

While all species were overrepresented, only four – buffalo, cheetah, elephant, and wild dog – were overrepresented in a manner that increased their proportion of the population of KNP. For all other species, their proportion in the population decreased as a result. This overrepresentation seems to be influenced by factors such as the abundance of the animal, the size of animal groups, the elusiveness of the animal, and the popularity of the species as presented through repost ratios, content type, and overall sighting and post counts.

*Table 8: Population abundance estimates for KNP compared to census estimates. Social Media estimates were achieved by implementing Lincoln-Peterson Index to account for instances of capture where n=(C*R)/Rm; n=population, C=sum of species quantity by sightings, R=sum of species quantity by posts, Rm=sightings count-posts count. Census estimates from past literature (Roocroft, 2021; Wray, 2023). %Social Media Estimate and %Census Estimate represent the proportion of that species' population within the park. Difference represents how closely Social Media Estimate corresponds with Census Estimate.*

Top Species	Social Media Estimate	%Social Media Estimate	Census Estimate	%Census Estimate	Difference
Antelope	599,440	65.53	232,820	70.37	x2.57
Buffalo	194,573	21.27	40,000	12.09	x4.86
Cheetah	2,234	0.24	150	0.05	x14.8
Elephant	38,979	4.26	13,000	3.93	x2.99
Giraffe	9,471	1.04	8,000	2.42	x1.18
Leopard	2,464	0.27	2,000	0.60	x1.23
Lion	3,655	0.40	1,600	0.48	x2.28
Spotted Hyena	5,026	0.55	3,000	0.91	x1.68
Wild Dog	11,635	1.27	300	0.09	x38.78
Zebra	47,239	5.16	30,000	9.07	x1.57

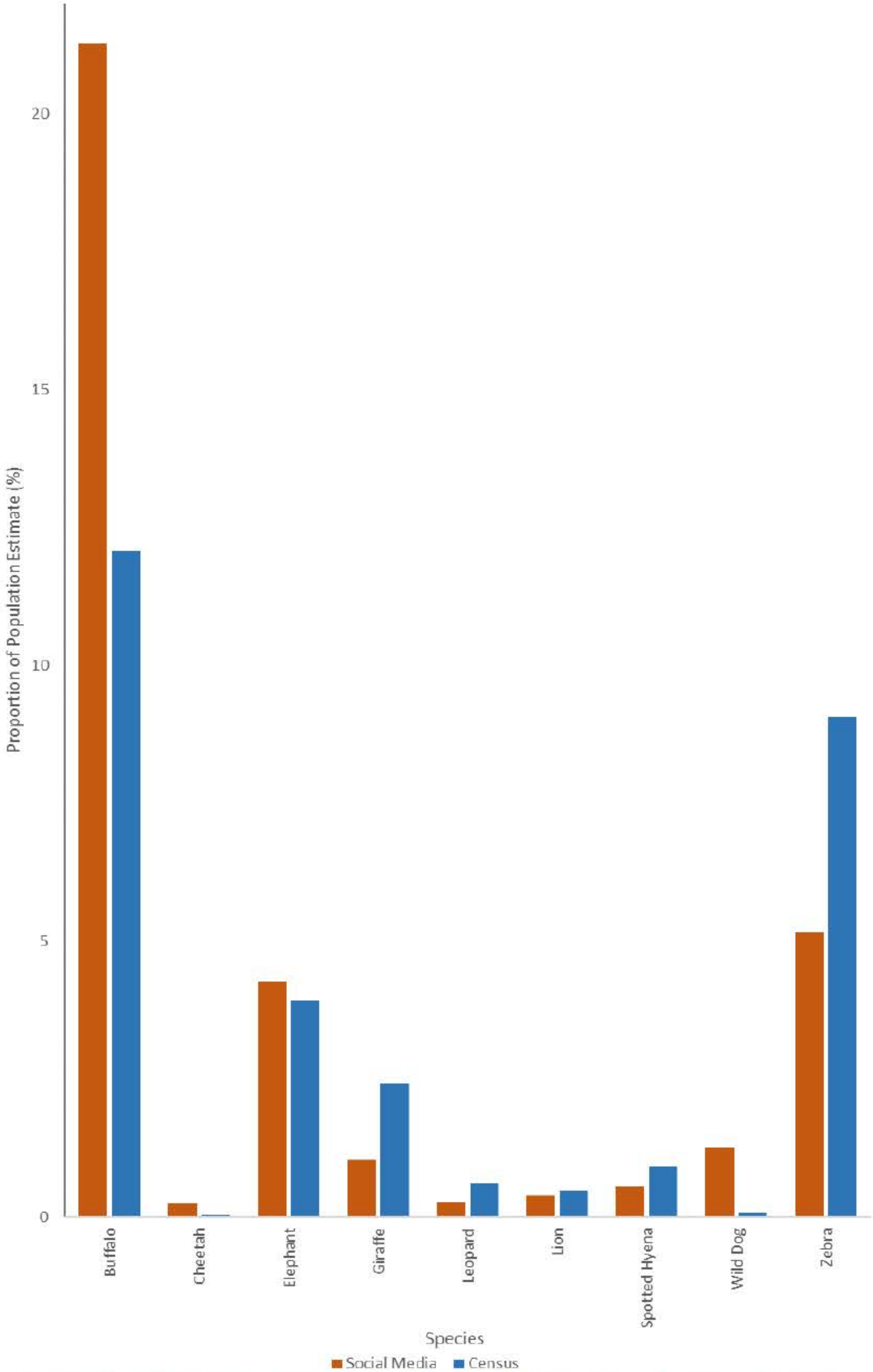

Figure 19: The proportional representation of species populations within KNP by source of abundance data estimate. Antelope contribution has been omitted to improve figure clarity.

4.3 User Habit Focus Groups

The focus group provided insights into the social media habits of ecologically minded individuals, focusing on three key areas: platform usage, subject matter, and content type. Findings are presented in figure 20. Most participants had Facebook, Instagram, WhatsApp, and Twitter accounts, but some were inactive on at least one of these platforms. Instagram and WhatsApp were most used by 19–23-year-olds, while Facebook was used less frequently and Twitter was primarily used by those aged 35+, with this being in a predominantly professional capacity.

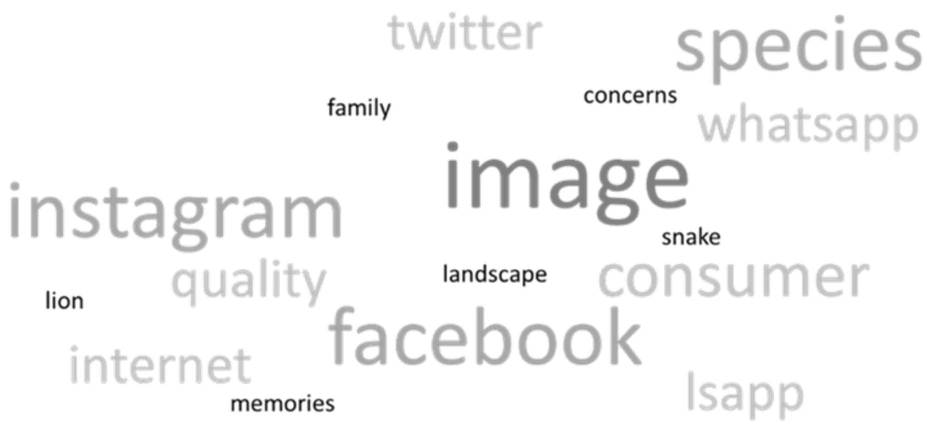

Figure 20: A depiction of the 15% most used key words referencing social media habits. Word size indicates frequency. For larger focus groups, participants' platform usage was tallied and incorporated into the transcript, ensuring consistent representation. Non-insightful "conversational/filler" words were excluded.

Across age groups and platforms, users tended to consume more content than they created. Those under 35 mentioned they would post to Instagram and Facebook after experiences like vacations or safaris. They intended to do so with collections of images. Participants without safari experience planned to post while abroad but could not, due to internet connectivity infrastructure. These same individuals also grew more appreciative of various, lesser-known species, and had not been aware of the dangers of sharing endangered species online but stated that they would act cautiously after their experiences. They emphasised that quality of (image) content would be a driver in selecting what they uploaded but noted that meaningful experiences or a sighting of a particular species would take precedence over quality. This last sentiment was shared by respondents with past safari experience, but their prioritisation of content would more highly regard new or rare experiences, and landscapes.

Lions were the top "bucket list" species for first-time safari goers, followed closely by snakes, towards which 19–23-year-old male respondents were more curious than other demographics. Based on responses, landscapes would be expected to be prominent uploads, with image being the primary content type expressing all subject matters. No users planned to include sighting date in their uploads and few had heard of LSApp before these focus groups but said that they would be likely to utilise it if they visited PNP or KNP (again) in the future and had access to a local SIM, but that they would likely not purchase a local SIM explicitly to use LSAPP.

Due to interactions with at-risk species, participants had become conscientious about the presence of location data in their posts and said that they would use generic location information such as country or province and would avoid uploading images with logos and other identifiers.

4.4 Signal Coverage and Latest Sightings Mobile Application Tings

Of the 39 attempts to retrieve mobile data services in PNP, just 17 had at least one device connect to a mobile network. In all cases where only one device connected, it was utilising a local Vodacom SIM. If there were two devices, it would be a local Vodacom SIM and a UK O2 SIM accessing MTN Fibre ISP, apart from the detection south of Lengau Dam where a UK EE SIM (MTN Fibre ISP) and UK GiffGaff (Telkom) device connected. Although 43.60% of PNP had apparent mobile data coverage, figure 21 presents that the distribution of this coverage is in concentrated pockets. These can be aptly described as pockets of no signal (0 devices connected), pockets of modest signal (1 or 2 devices connected), and pockets of strong signal (3 or 4 devices connected). 7.70% of sites had only one device receive service; 10.25% two devices (meaning 17.95% modest signal coverage); 7.70% of sites 3 devices, and 17.95% all devices connected (meaning 25.65% strong signal coverage). Still, the majority of PNP (56.40%) had no signal coverage.

As can be seen in figure 21, there appears to be no correlation between physical infrastructure and the likelihood of a device receiving signal. Artificial hides and the Pilanesberg centre did not have signal, despite the presence of built structures. There is also no increased likelihood of signal at any of the dams tested. Instead, there appears to be two factors which dictate signal coverage: proximity to the perimeter of PNP and altitude. There were two points near Lenong View, one of the highest points in PNP, where all four devices retrieved service, and a point on the route to this viewpoint where two devices connected. This area was otherwise a pocket of no signal. Otherwise, the majority of test points where signal was retrieved (82.35%) were located on Nkakane Drive near the eastern Manyane gate, or on the southern stretch of Kubu Road near Bakubung Gate.

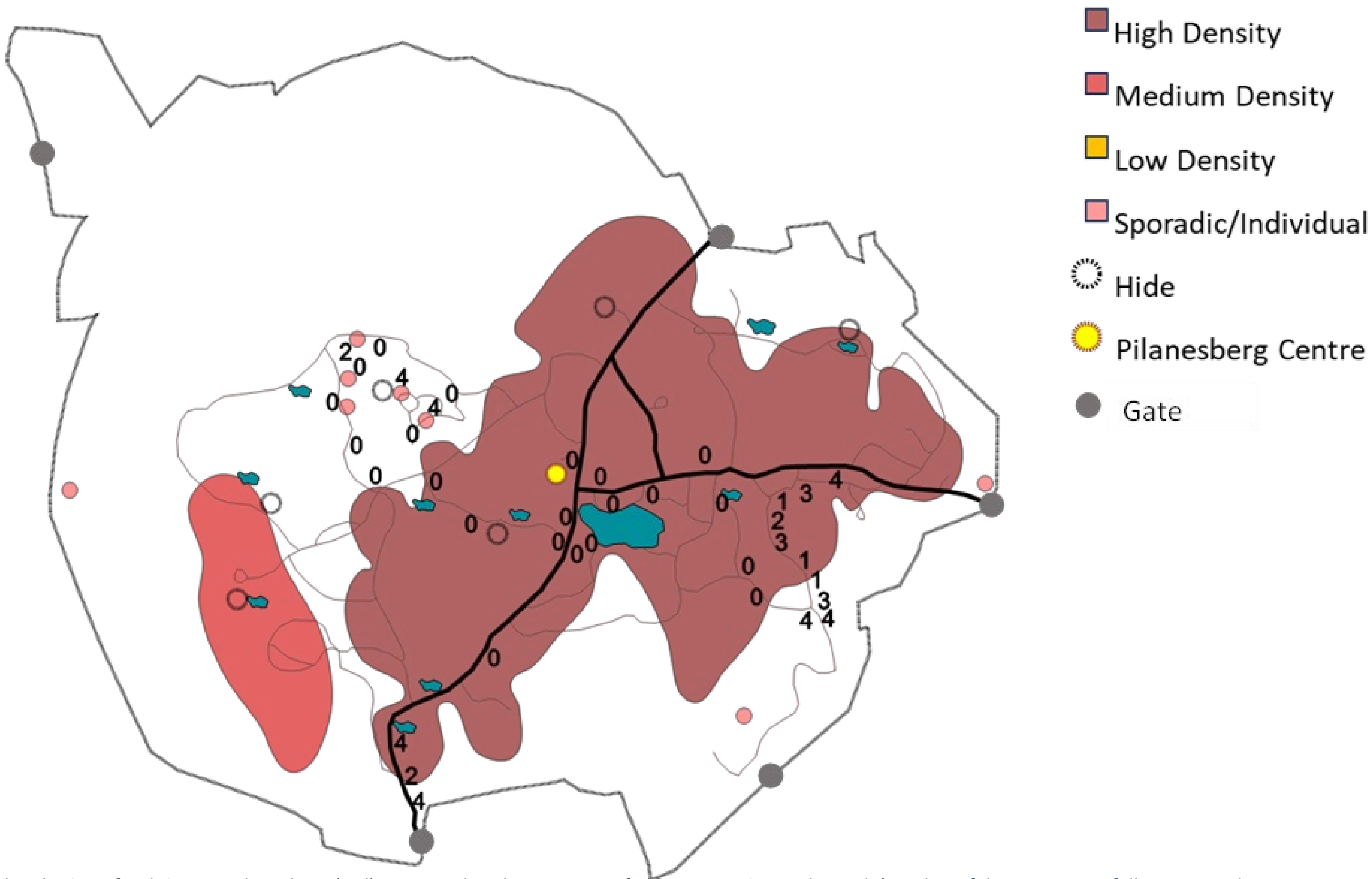

Figure 21: The distribution of sightings in Pilanesberg (red) compared to the coverage of phone reception in the park (number of devices successfully connected presented as black numbers).

The majority of mammal sightings occurred within the dark read area highlighted in figure 21. The sightings here were so densely populated that distinguishing specific points of interest within the matrix was impossible. The majority of sightings are concentrated in the centre of the park and towards the eastern gates, along Kgabo and Tshwene; the two main roads of PNP. The dirt tracks (drives) off of Kgabo and Tshwene also possessed a high concentration of sightings, though this was mostly among drives which were short and looped back to a main road relatively quickly. The medium density cluster on the west of the figure were predominantly taxa which exhibited high frequencies of locomotive behaviour; elephant, giraffe, and antelope, whereas the three sporadic/individual sightings indicated towards the perimeter of the park were all lion sightings. The individual/sporadic sightings by Lenong view (north-west of the main cluster) were smaller species such as rock hyrax (dassie) and steenbok. These sightings do somewhat correspond to the distribution of phone signal. There are individual sightings around Lenong view, where there were pockets of strong signal. The majority of signal coverage test points were in the main matrix of sightings, meaning that the majority of test sites with modest signal coverage coincidentally fall within these confines. The two do not seem to be causal, however, as the majority of pockets of strong signal fall beyond the boundaries of the high-density sightings.

Section 5: Discussion

5.1 Summary of Findings

This study found that there is an abundance of conservation-relevant data present as a by-product of other user behaviours on social media platforms. Although the immense volume of data on LSApp shows the advantage of having a dedicated application for a cause, social media platforms still contained lots of data. It was found that the abundance of posts on platforms were dictated by the activity levels of users within the parks, with KNP visitors seemingly five times more active on social media than PNP visitors. This activity seemed to be dictated by the different popularity of platforms among park goers, which was in turn dictated by the types of people visiting the parks.

These posts were sufficient to obtain a species inventory of 55 mammal species. LSApp, Facebook, and Flickr were the only platforms to detect completely unique species, indicating that (apart from WhatsApp), studying any single platform would be sufficient to create a comprehensive species inventory. Again, in detecting 81.82% of all species as sightings, LSApp did still outperform social media platforms; but when observing posts, Twitter possessed the most thorough species coverage, with post from Twitter containing species-specific data for 89.09% of detected species (49 species). This study did, however, find that the representation of species varied substantial across platforms. This (among other factors) raises difficulties in attempts to monitor abundance and population from these data. Post frequency varied by species and by platform, but the same species were consistently popular, just to significantly varying extents. Lion, for example, were five times as frequently sighted on LSApp as other platforms.

5.2 Overview

Over the 114-day study period, the number of daily posts from platforms varied. These variations are analogous of two variables; (1) park visitor numbers; (2) social media user activity (Tankenen, *et al*., 2017). The sociodemographic composition of a park's visitors alongside varying visitor rates has led to tangible impacts on the distribution of posts (also witnessed by Tankenen, *et al*., 2017). Firstly, as presented in figure 4, the two peaks in post counts correspond with the Easter period and South Africa Freedom Day. Both are holidays celebrated in South Africa (Marschall, 2014), and with Easter being a global celebration, public holidays and school breaks often correspond to the occasion.

Increased post numbers around these dates are likely caused by the presence of these holidays, which themselves provide insight into visitor compositions. PNP receives a larger share of its visitors from domestic tourism (Nikodinoska, *et al*., 2014) than KNP; likely contributing to the increased growth in post numbers in PNP on day 73 (just after South Africa Freedom Day). KNP is unlikely to experience the same growth due to its increased patronage by international tourists increasing prices, and its increased competition from other nearby reserves and attractions. KNP is also significantly larger than PNP (figure 1), and while PNP can be "completed" in a day, KNP cannot, potentially reducing the experience.

Although the origins of tourists may explain fluctuations in post rates on these particular days, the increased activity of guests in KNP is likely caused by several factors. KNP receives 300% the annual visitors of PNP, but sightings were 600% as prevalent. Even considering the increased species richness of KNP, OCs in the park still seemingly post 70% more frequently than those within PNP. This is likely caused by the composition of platforms in the parks and the specific accounts using them, discussed later, a series of phenomena which would contribute to increased user activity at parks with higher visitor counts (Tankenen, *et al*., 2017).

5.3 Species Representation in online spaces

Of the 8,455 sightings taken as data, LSApp provided the greatest share (45%) and accounting for reposts, it provided 69.34% of sighting and post data. The sheer volume of posts is likely causal to LSApp's rapid accrual of species, its vast unique species coverage, and its influence over other platforms either directly, through repost data, or indirectly, by OCs adopting LSApp features and formatting in their own posts on other platforms. Despite this volume of sightings and the rate of propagation of LSApp posts, due to additional posts made uniquely to those platforms; WhatsApp and Twitter ultimately present more post data, and when assessing posts, the addition of bushbuck (*Tragelaphus scriptus*), black rhinoceros (*D. bicornis*), southern white rhinoceros (*C. simum simum*), and tree squirrel (*Paraxerus cepapi*) to LSApp's already extensive species library makes Twitter possess the greatest species inventory of any platform studied.

When observing sightings, LSApp, Instagram, Facebook, and Flickr all created species inventories that compare to and exceed the capabilities of traditional biodiversity survey methods. Janecke (2021) conducted a species richness survey at catenae and nearby waterholes in KNP. They recorded 33 mammal species (of the size of a mongoose or larger as smaller species may not have been captured) via targeted camera traps. This manual retrieval methodology detected 53 mammal species in KNP (47 of which were equivalent in size to mongoose or larger), and 31 of the species detected by Janecke (2021).

Among the species detected through use and application of this methodology, 18 of the 22 species within the Bovidae family believed to be extant in KNP (10 more than Janecke detected) (SANParks, 2007) were seen, alongside all of the species which could be considered "iconic" African big game (lion, leopard, elephant, buffalo, giraffe, and hippo), as well as pangolin and both extant rhinoceros species. This large array of diverse species would serve to create a varied and comprehensive species inventory, though one may have to delve through a large number of lion posts before encountering one about an elephant shrew.

The repost and multipost rates of species predominantly corresponded to the activities and preferences of business accounts; the most prevalent source of reposts among any account type. Due to the motivations of businesses, often to advertise their product or experience in an online space, they select the same species and content to repost: leopard, lion, wild dog, cheetah, elephant, and buffalo (as presented in table 6) (Curtin, 2010). All of these species are iconic, charismatic, and widely popular (as is evident by the number of sightings and posts relative to their population sizes). Ultimately the drivers behind a repost being made were the same as those behind a sighting being recorded by an OC at all. Subsequently insights into platform usage and account type (discussed later on) also apply to reposts and multiposts.

5.3.1 Platforms as Environments

The Bray-Curtis dissimilarity scores obtained in this study prove that species representation is varied across social media platforms and applications. Individual taxa are represented diversely across platforms, and platforms represent unique compositions of the most prevalent species. These trends are ultimately caused by how and why people use social media in the ways that they do (Tankenen, *et al.*, 2017). LSApp is unique in having a dedicated and prescribed use case since inception, but traditional social media platforms, such as Twitter, Instagram, and Facebook, have developed unique identities, communities, and use-cases over years of use by millions of users (Gruzd and Haythornthwaite, 2013).

Twitter, Instagram, and Facebook possessed the lowest Bray-Curtis dissimilarity scores, and were more dissimilar to LSApp (from which a significantly larger number of sightings were observed), and Flickr and WhatsApp (from which a significantly smaller number of sightings were observed). Where Bray-Curtis treats sightings as population counts, the dissimilarity scores reflect differing detection (and recording and uploading) rates across platforms, rather than species presence within the parks themselves. Although this could be considered analogous, it is more a measure of data availability and variability than it is for species composition (Baselga, 2013). The Chi-square assessment of association between platform and species composition as much justifies species distribution as it does the diversity of data across platforms. This diversity is not caused by the platforms themselves, but rather through the interactions between users (account type), their preferences of subject matter (species and behaviour), and the content that they consume and/or create (Tankenen, *et al.*, 2017). The PERMANOVA contract obtained in the same section confirms that the same species is represented variably across platforms (habitats) (Kelly, *et al.*, 2015), indicating that so too do all drivers of a sighting's creation.

5.3.2 Species Accessibility across Platforms

Species accessibility, akin to effort spent in other studies (Hart, *et al.*, 2022), varied across platforms. Facebook required the least effort to accumulate a moderate number of species and would go on to detect a significant number of unique species. It was the fastest platform to detect all of the most prevalent taxa and would subsequently be a viable source of rapid confirmation of a species extant status in a park. Facebook did, however, fail to cover many of the “less popular” species, with over one-third of species sighted on the platform belonging to the Bovidae family. In order to study the less prominent (popular or abundant) species within these parks, LSApp (and to a lesser extent Flickr (Edwards, *et al*., 2021)) should be employed. LSApp presented the second fastest rate of species accrual and the most complete species inventory of any platform (apart from Twitter when considering posts). Many of LSApp’s detections of these less popular species, however, were later into the observation period, and they were only sighted once, so researchers pursuing time-sensitive confirmation of a species would ultimately be relying on luck. It is possible on LSApp (but easier on their WhatsApp group chats) to review a historical catalogue of sightings from the timeline, so historical presence could be confirmed in this manner.

Otherwise, species accumulation rates are superior across most platforms to traditional survey methods (over the first 10 days or equivalent effort). Hart, *et al*. (2022) report that after 250 trapping nights at a reserve in South Africa, 15 species were discovered by targeted camera traps, and that over 300 nights, 12 were discovered by random camera traps. Comparatively, social media posts present this same number of captures in an average of 11 days, and in half as much time sight almost four times as many species.

This illustrates that not only do observations from online mammal sightings have great potential to construct extensive species inventories, but that they do so with considerably less effort than existing survey approaches (Garrard, *et al*., 2008). The ability to refer to past records of data on these platforms, although easier on some than others, affords the ability for a researcher to cover days or months of species records in a matter of hours if they were so inclined. This capacity exceeds that of other citizen science studies too, where Dunn and Winkler (1999) accumulated 3,450 data points for 40 years’ of nest data, while the methodology outlined within this paper observed four times as many in less than half a year. This stands as a testament to the potential of social media data not only for accessibility but also volume of data (Deng, *et al*., 2015; Edwards, *et al*., 2021; Jarić, *et al*., 2021).

5.3.3 Social Media Estimates of Species Abundance

Abundance estimates are the first hurdle for this data source (Pech, *et al*. 2004). Due to the extensive reposting, recapture, and overrepresentation of animals, a study such as this which focused on volume and breadth of data could not confidently track the recapture of animal sightings (Madon, *et al*., 2011). On an hourly, daily, and monthly scale, the same individuals would be unknowingly captured by multiple tourists, and uploaded in a manner that did not meet the criteria to be defined as a repost or multipost, or which were even identified as repeated sightings in the first place. There were some cases, such as for an injured cheetah who received surgery and was recovering, where social media users (in this case those on LSApp and Facebook) would actively seek out sightings of the animal to keep their communities updated. Again, in many cases these posts did not meet the requirements to be classified as reposts or multiposts.

The increased amount of content for certain species (at the expense of others) leads to abundance estimates from this dataset as it currently stands being completely inaccurate. Although the vast majority of sightings and posts possessed count data which was verified where possible, during data collection, the volume of recaptures of popular, charismatic animals lead to inflated numbers. In order to obtain an accurate population estimate, individuals and/or groups/herds of animals would have to be systematically identified to account for rates of recapture. Although it was not a measured metric, during the data collection it was noted that it was possible to track an individual or small group of animals for days at a time due to high sighting frequencies and relatively consistent locations. A coalition of three male lions in PNP, for example, was able to be tracked with confidence for nine days due to the distinctive facial scarring seen on one member of the group, and their tendency to remain constant in their movement. Similar markings and patterns could be used to identify animals that live in small groups (Madon, *et al*., 2011), but for larger herds such as those of impala and buffalo, or fluid herds such as giraffe, greater levels of care would need to be taken. Although giraffe have unique marking patterns, the quality of images or videos, and the lack of visual representation in text posts would lead to increased uncertainty in identifying individuals consistently. Many animals in the parks have tags or collars as a result of relocation, research projects, or to monitor their health after medical intervention, which could act as easily identifiable features (Madon, *et al*., 2011) but again, there are many potential sources of uncertainty.

Correlations between species and overrepresentation are likely caused by similar phenomenon to those between species and repost rate, ultimately being determined not only by the likelihood of an animal being sighted, but by the popularity of that species (Mangachena and Pickering, 2021) and the potential application of its content to marketing (resulting in increased tour company sightings or business account reposts), and how rare or prestigious a sighting is perceived (Van der Merwe, 2017).

Ultimately, the measures taken to account for recapture of individuals in this study proved inadequate to provide an accurate representation of species abundance in KNP, but there are actions that can be taken to remedy this.

5.4 Park Features impacting Mammal Sightings

The sociodemographic compositions of park visitors (touched on in section 5.1) were not limited to the origin of a tourist. Park patronage can be determined by several factors, including income, age, and reason for visiting (Tankenen, *et al*., 2017). These in turn represent subtleties within account type data that can be further investigated.

KNP is a world-renowned safari destination and has a booming tourism industry as a result. The park alone hosts dozens of holiday lodges ranging from cost-friendly accommodation to luxury, spa-inclusive experiences (Siyabona Africa, 2023). KNP garners this reputation from its history of wildlife preservation, its size, and its ability to fully immerse guests in a safari experience. All combined, the park is seen as an expensive, luxury experience, even if OCs opt for the most cost-friendly options.

Online interpersonal influence marketing is a growing phenomenon on social media that sees "influencers" (Hudders, *et al*., 2021) display their lives in luxurious and romanticised manners to amass followers to whom they market products in behaviours often called "peacocking" (Creevey, *et al*., 2019; Rotman, *et al*., 2021; Leung, *et al*., 2022). These influencers often utilise international travel to achieve these ends. For many such individuals, KNP would likely be a preferable destination to PNP due to its widely known name contributing to publicity efforts. Pilanesberg, on the other hand, and as already stated, receives 82% of its patrons as domestic day visitors (Nikodinoska, *et al*., 2014). These factors in turn correspond to ages (with PNP's average visitor ageing between 25-35 years of age).

These features of park visitors correspond to features of social media users. Age, income, and occupation all correlate to what social media platforms people use (Sloan, *et al*., 2015; Ozimek and Bierhoff, 2016; Pittman and Reich; 2016), and as such the visitors of a park are essentially represented by the account types of users referencing them (Tankenen, *et al*., 2017). This likely contributes to the slight increase in tour company prevalence in Pilanesberg, as guides have to be more active in online spaces to capture the audience of potential clients.

Furthermore, geographical features play a role in the distribution, visibility, and behaviours of animals. PNP has significantly more varied elevations than KNP, and many peaks and slopes within the park are inaccessible to visitors, but not to animals. KNP, on the other hand, is relatively flat with low topographical variation, and would as such experience increased visibility, but is so vast that a single (group of) visitors would likely only cover a portion of the park in detail. Both of these features contribute to different rates of capture and likely towards the prevalence of species.

Ultimately, the increased capacity, popularity, visibility, and species abundance of KNP appear to have accumulated to an increased number of sightings within the park. Though, considering the numbers posed in section 4.1, one would expect to see twice as many posts from KNP as PNP (based on the difference between visitor numbers and sighting counts (Lacea and Ferreira, 2014, Ferreira and Harmse, 2014)); though this modifier is closer to 1.85 times the amount. This implies that users engaging with content from PNP are considerably more active than those from KNP in reference to propagating and multiposting sightings.

5.5 Causes for and Impact of Post Lag

Although visitor numbers appear to be a causal factor to post volume (Tankenen, *et al*., 2017), they do not appear to be causal to post lag. The only correlation post lag presented with any other data type was platform, though this is caused by the absence of sighting date data on any platform other than LSApp or Flickr. Post lag was reduced significantly by the data from LSApp, where the overnight delay of an evening sighting being reposted to Twitter or WhatsApp the next morning was the source of the majority of LSApp's contribution to the data. Post lag from sightings on Flickr were significantly greater, likely due to the lack of compression when uploading on the platform (and the potential for users to improve the quality of their image through editing) meaning that OCs wait until they return from holiday to upload. Additionally, due to the use of dedicated cameras (opposed to those on their smartphones), Flickr users require additional equipment (a computer or laptop) to transfer their images to a format and device which can suitably upload the content. There are now cameras on the market with built in image transfer and upload capabilities, but these features still rely on a strong and stable internet connection, and cameras with this feature would not yet be considered the norm.

The lack of fast and reliable internet connection limits timely uploads of mammal sightings on all platforms. The signal coverage tests conducted according to section 3.15 and section 4.4 indicate how prevalent the retrieval of signal was at PNP but fail to consider the actual application of this metric.

The average bandwidth (the theoretical maximum transfer load of a connection between a client and server) between KNP and Twitter's Sacramento server leads to a response time of 578.75ms for a 32Byte packet of data. This 32Byte packet converts to 256bits, leading to a bitrate of 2,260kbps between KNP and Twitter. This is ultimately half the recommended bitrate of 5,000kbps as prescribed by Twitter for interacting with video posts (X, 2023). A wildlife sighting image taken on a mobile phone had an average file size of 2.995MB, meaning that it would take approximately three hours to upload, provided that a stable mobile service of moderately high strength (4G connection) can be maintained for that period of time.

This demand for bandwidth ultimately leads to a lack of real time posting by OCs. LSApp is the only platform in this study which actively encourages wildlife sightings to be current and relevant to that sighting day (Latest Sightings, 2023). Twitter's content generally also has a current and time-sensitive theme, but where the infrastructure does not support it, Twitter users likely upload their sightings soon after their return home, or at their accommodation if its Wi-Fi facilities can support posts.

Post lag ultimately leads to increased uncertainties in data. Sightings and posts without a stated post lag could have been missed as repost or multiposts if the time frame of the post lag were significant enough that the manual detection of additional posts did not identify it. Furthermore, it reduces the relevance of the information in the post itself. Location data and behaviour will almost certainly have changed after a matter of hours, and for species such as giraffe which have particularly fluid herd structures (Bercovitch and Berry, 2013), animal counts will even have changed.

5.6 Data Types: Account Types

Here, account type could be used as an interpretation of the motivation of users behind creating sightings and propagating posts. Ultimately, the groups chosen reflect individuals whose accounts would have assumed behaviours and reasons to post in the manner they did. Business accounts, for example, are advertising a good or service. Holiday companies are advertising safari destinations, resorts and camps are advertising their accommodation options, and even the Latest Sightings team are attempting to receive patrons to LSApp and advertisers to support their development and projects. Due to these motivations, business accounts would upload and repost content which they believe is going to appeal most to their target audience; curating their content to the most popular species (in this case lion, leopard, and elephant) and to specific content types which can achieve this effectively; image and video (figure 12).

Latest Sightings is the exception here. LSApp reposts propagated to Twitter and WhatsApp were classed as business posts as the Latest Sightings team was responsible for the act of reposting. Subsequently, their content on these platforms showed no such curation. The Latest Sightings team were also relatively inactive on Instagram, or their posts were not revealed by the search functions of the platform and had no presence on Flickr. Their behaviour mostly deviated from the norm on Facebook, where they were perpetrators of the behaviour which will be discussed in section 5.12 regarding the subject matter of long-form video posts.

Non-Latest Sightings business accounts, which were primarily active on Instagram, appeared to use their sightings and posts to advertise the beauty of the natural world and the potentially amazing experiences visitors could have with wildlife (Creevey, *et al*., 2019). As such, they utilised image portrait posts more than any other account type. It is likely that the businesses themselves did not make these sightings but rather purchased the images for their marketing uses. As such, there are many sources of unreliability among these data; namely the likely presence of post lag which has no means of being quantified, and the chance that these accounts cycle through content periodically, increasing the likelihood of unaccounted for recaptures.

Tour companies operated under similar motivations to business accounts; their sightings and posts advertised a product (Curtin, 2010). Sightings from accounts that could confidently be identified as tour companies were scarcely identified on any platform other than LSApp (where their names were italicised with personal account names remaining in standard font, though this formatting seems to have been removed since data collection); though due to this, posts from tour companies were propagated to Twitter and WhatsApp. Their concentrated activity on LSApp is likely the result of three observations that these users (and this study) made: (1) LSApp's format allows for uploads under signal and service conditions that other platforms do not support; (2) LSApp is widely monitored by users who will be visiting the parks imminently, making them a reliable potential customer; (3) the Latest Sightings team propagates their sightings to Twitter and WhatsApp on their behalf, fully crediting the creator of the sighting and thus advertising their services.

Although uploading sightings acts as an effort effective and cheap way to advertise their services, it is likely that tour companies are wary of the effect that uploading certain content types and how they advertise their services has on their business (Margatyan and Wall-Reinius, 2017). Tour companies exhibited the greatest share of their sightings as text posts. The majority of these text posts were of lions, leopards, and cheetah (see figure A6), with wild dog also being popular among the account type, and the behaviours within these posts disproportionately represented resting/stationery and locomotion, with all other behaviour types being represented far less than from business or personal accounts.

This format of posting seems to have been conducted to protect the intellectual property of the sighting itself from other tour companies (so that the sighting site did not become overcrowded, and the quality of experience diminished). It is also likely that, in correlation with this, posts lack visual content to (1) reduce the infrastructure requirements of uploading the sighting; (2) conceal the fact that a sighting has low visibility or is of poor quality from their potential clients; (3) prevent other operators or guests from recreating the sighting using landmarks present in the image; (4) allow for the fabrication of a sighting altogether (Blenkinsop, 2018). The evidence for any and all of these possibilities is circumstantial, and an investigation into the operations of tour companies as the primary subject would need to be complete to ensure that the marketing motivations behind their sightings did not impact the quality of information that they contained.

It therefore seems that personal accounts might present the least bias to any species, behaviour, content type, or platform. This includes all potential influencer accounts which were classified as personal accounts. Although an influencer's motivations lie in advertising, their posting structure is built on interpersonal marketing (Pittman and Reich, 2016), which relies on their appearance as familiar and friendly with their followers (Tafesse and Wood, 2021). As such, it can be assumed that influencers base their posting habits on the preferences of their followers which would in turn represent similar content patterns to typical personal accounts. The only exception here is how content was posted, where accounts that appeared to be influencers would utilise collections (typically as a reconstruction of their day) where some of the images did not include wildlife sightings, but rather recreated their day's experiences as a whole including meals and accommodation. As such, influencer collections provide a lower concentration of mammal sighting information, though there was no apparent reduction in quality of post.

Personal account sightings presented the lowest rate of overrepresentation of lion, leopard, and cheetah. Antelope (a group which possesses a large amount of diversity as defined in section 3.7), giraffe and zebra, which all have relatively large populations within the parks, are represented significantly more by personal account sightings than either other account type. Table 9 briefly presents how representation of taxa by personal accounts corresponds closer to population census estimates in KNP than business or tour company accounts. In this table, personal accounts deviated on average by 2.8 places, business accounts by 4.6, and tour companies by 4.8.

Although lion was still the most popular species for personal accounts, it was interesting to observe a slight preference towards elephants compared to other account types. Cheetah and wild dog are represented by personal accounts much more proportionally to their populations, indicating that abundance plays a greater role in the detection of these species and subsequently their sightings by personal accounts. This is likely contributed to by the increased number of novice visitors traversing the parks without tour companies or guides making less successful sightings within this account type class. Strangely, buffalo were as underrepresented by personal accounts as other account types; though this may perhaps be due to their abundance making sightings less significant, or the fact that buffalos might not be particularly popular due to their appearance or wallowing habits (Mangachena and Pickering, 2021).

*Table 9: Hierarchy of taxa ranked according to census population estimates (1=most populous, 10=least populous, which is also how taxa are ordered), and by popularity of taxa sighted by each account type (1=most sightings from that account type, 10=least sightings from that account type). Cells are shaded based on their relative position in this hierarchy, with the gradient presented in Census Population Estimate Ranking column being the control; if an account type's cell shade is similar to the shade in the Census Population Estimate Ranking column, this indicates that their representation by that account is proportional to the species' abundance. *Zebra and Giraffe experienced the same number of sightings reported by Tour Companies and are as such both ranked 9.*

Species	Census Population Estimate	Business Account	Personal Account	Tour Company Account
Antelope	1	10	4	8
Buffalo	2	7	7	7
Zebra	3	9	8	9*
Elephant	4	3	2	5
Giraffe	5	8	5	9*
Spotted Hyena	6	4	6	6
Leopard	7	2	3	2
Lion	8	1	1	1
Wild Dog	9	5	10	4
Cheetah	10	6	9	3

Antelope were much more proportionally represented by personal accounts than business or tour company accounts. The antelope group contains 15 antelope species (as defined in section 3.7), and as such presents greater species diversity than all other top species combined. The antelope group contains particularly abundant species such as impala and wildebeest, but also uncommon, rare, and wonderful species such as sable and bushbuck who can grow impressive sets of horns, klipspringer and duiker whose small stature likely increases their perception as timid or playful (Mangachena and Pickering, 2021), and tsessebe and springbok who are among the fastest antelope species on the planet (with the springbok having added cultural significance as the mascot of South Africa's national rugby union team).

This diversity means that a large amount of the time, for personal accounts which were mostly run by people for whom a safari was an engaging and exciting experience, sighting an antelope meant sighting a new species. The focus groups conducted in May 2023 confirmed that the excitement of a new experience played a considerable role in the likelihood of a piece of content being uploaded. Where many tourists might not appreciate the diversity of the antelope species group, each new species discovery will likely subvert their expectations and subsequently increase their likelihood of posting.

Among focus group respondents who had experienced safaris numerous times before, many stated that there were still species of antelope supposedly extant in the modest reserve they were visiting and/or PNP that they had not seen, even after half a dozen or more excursions. Subsequently, a sighting of this species may be even more significant to them than a repeated sighting of a lion or leopard, and therefore take precedence over such sightings when deciding which piece of content to upload.

This is likely the ultimate factor in this discrepancy of representation. For the intentions of sightings uploads by business and tour company accounts, the subject matter needs to be an instantly recognisable, broadly adored, and iconic species. Here, lions and leopards and elephants act as mascots for their livelihoods, and the potential for a post of a nyala or eland, for example, to be overlooked by a potential client as "just another antelope" would cost them money and time. Ultimately, there is a personal disconnect between these accounts and their content, where they must rely on the promise of potential to make a sighting worth uploading (Curtin, 2010). For a personal account, every sighting uploaded will be accompanied by a memory of the experience: the neophilia associated with seeing an iconic species in the wild for the first time or discovering a new favourite they previously did not know existed.

5.7 Data Types: Content Type

Content type varied by species, platform, and account type, with particular subject matters being captured in certain manners, alongside platforms that have features and communities with preferences regarding content shown, and the motivations for a user's post all contribute to these variations.

Tour Company accounts utilise text posts the most broadly, as discussed in section 5.6, followed by business accounts, whereby informing potential clients of the existence of a sighting while not devaluing that experience by relying on visual content better serves their intentions. The majority of sightings without behaviours were subsequently text posts from tour company accounts. These posts are also more effort conservative to construct and upload where signal infrastructure might be limited in the field, or where they might want to move on rapidly to the next sighting to improve the experience of their existing customers. Personal accounts may resort to text posts where a sighting was too short or immersive for the OC to capture it as an image or video (Tankenen, *et al*., 2017).

While some species and behaviours might be prohibitive of certain content types in this context, others seem so significant that they had to be visually captured. Instances of intraspecies interaction, interspecies interaction, and mating were predominantly captured as images. The portrait behaviour type was exclusively visual content, due to the nature of the definition of the behaviour. Other behaviours such as resting/stationery, locomotion, or feeding were captured most by images simply due to this being the most prevalent content type, though the ease of taking an image over recording a video likely played a role in the decision making.

Due to the issues of intermittent signal coverage within the parks and the demands that uploading large files (images and videos) impose, many users may withhold sightings from platforms that encourage relevant and current posting such as LSApp and Twitter, and instead opt to upload their experiences to Instagram and Facebook at a later date. In this way, platform can be consequential to content type, though in some cases it can be causal. Users who only post on Instagram or Flickr, for example, would likely only record sightings as images or videos. In the case of the latter, sightings were only recorded from content taken on a dedicated camera device, which would impose greater difficulty to the upload process, and was proven to introduce long periods of post lag. Subsequently, intended platform not only determined how a sighting was recorded, but also when it was uploaded.

Long-form video sightings were almost exclusively generated by personal accounts where an individual can determine how long they want to spend with a sighting, whereas businesses or tour companies may move on in hopes of providing their client with a "better" sighting of a species they deem preferable (lion, leopard, cheetah). The subject matter of these videos, often long instances of predators hunting or of intraspecies interaction, are however appreciated by business accounts who propagate them across platforms as reposts. This implies that the practical limitations of capturing and uploading a post as certain content types is often prohibitive to their creation.

5.8 Data Types: Multiposts and Reposts

The drivers of additional posts have been considered throughout the analyses of general trends, account type, and content type. It appears that although sightings might be influenced by neophilia – the enjoyment of discovering new knowledge/experiences – reposting is less subjective. The memory attached to seeing an antelope species that an OC had not previously known of is a driver for a sighting to be uploaded, but for other platform users who do not share the same memories and sentiment, the propagation of such a post likely has little appeal. This also likely influenced by user habits dictated by their self-esteem (as discussed by Greitemeyer (2016)'s insights into social media users). Even if an individual were to appreciate the experience an OC had while recording an obscure species for the first time, they may not share the sighting due to fear that their followers would not find it relevant. Instead, they would be more likely to interact with (like) the post which would have an indirect effect on its propagation but would not create another instance of the post.

Otherwise, multiposts and reposts are driven by the accounts themselves. The vast majority of reposts come from LSApp and are reposted by the Latest Sightings team, a business account. Their intentions for doing this are similar to the intentions of other business accounts posting sightings; to advertise their product. Reposts by businesses other than Latest Sightings would essentially serve either to advertise another business/tour company or show that independent excursions can provide high-quality sighting experiences, defeating the purpose of advertising their service. There were no instances of tour companies reposting, likely for this very reason, and under similar logic multiposts would serve to devalue the experience to potential customers and were not widely observed from business or tour company accounts as such.

Content type seemed to drive repost rates, as did the platform they were recorded on. Images and videos are more engaging on a timeline and as such propagating these sightings will lead to higher engagement with the shared content (Shahbaznezhad, *et al*., 2022). Twitter and WhatsApp received reposts from LSApp due to the Twitter users' emphasis on recent content, and the curated and dedicated use case of the WhatsApp group chats. Instagram and Flickr do not support reposts or sharing in a manner conducive to data collection, but there were instances on Instagram of multiposting where multiple members from a party all posted their perspectives of a sighting, or one individual who uploaded every piece of content they created three times from three unique accounts (identified due to the posts containing the same watermark which contained a name that was present/referred to in some format in the display names of all of these accounts).

5.9 Data Types: Species

Species may be the most prominent driving factor in the creation of a sighting (Wood, *et al*., 2013; Hausmann, *et al*., 2017). The desire of an OC to sight a species will inform their pursuits, the routes of their excursions through the park(s), and the content type they wish to collect (similar to how they may wish to sight certain species; Di Minin,*et al*, 2013; Hausmann, *et al*., 2017). The ability of an OC to capture said sighting is determined by the abundance and behaviours of the species, as well as a considerable amount of luck (which is itself an expression of many unquantifiable external factors such as weather, the location of the animal relevant to the OC, the level of disturbance other visitors have created that might have driven animals to avoid roads).

The drivers behind an OC's desire to sight a species are subjective, but observations from the focus groups and the data collected suggest that lion meet all these criteria (Willemen,*et al*., 2015). Elephants and wild dogs share a perceived emotional intelligence and strong familial structures with lions, whereas both lions and leopards are carnivores and witnessing a successful hunt (or simply a creature so capable of hunting) seems to fulfil a morbid curiosity among OCs similar to that typically expressed by dark tourism (Krostanje, 2020; Mangachena and Pickering, 2021), but perhaps less niche. The striking manes of lions, fur patterns of leopard, cheetah, zebra, wild dog and giraffe seem to drive OC interest (Mangachena and Pickering, 2021), whereas the overall elusiveness of leopard and cheetah would make sightings feel more significant. Ultimately, quantifying any of these drivers would require an in-depth analysis of visitor preferences through interviews and focus groups, but doing so would potentially allow for a statistical consideration of these biases and their contribution to the overrepresentation of particular species, allowing for this overrepresentation to be mitigated in abundance estimates.

The focus groups completed here provided some insight into this phenomenon, but not enough to quantify. There appeared to be a level of expectation management when individuals responded to questions about what species they were looking forward to seeing. Rhino and giraffe were at the top of the list, likely informed by past excursions to the wildlife reserve having high success rates in sighting the species, but also lions and elephants during their holiday day to PNP. Very few respondents listed cheetah or leopard highly on their list, despite their presence within PNP and the presence of leopard within the reserve they were visiting. This seems to have been through a mutual but unspoken attempt to avoid disappointment if such a sighting did not come to fruition, as the response after successfully capturing a leopard on a camera trap was observed to be of equivalent or greater intensity to their successful in-person sighting of lion in PNP. This expectation management might ultimately serve to increase the number of leopard sightings that are uploaded, as the excitement of sighting one increases the motivation to post it, whereas for species OCs "expected" to see (or had higher hopes of sighting), this drive is lesser. These observations align with those of Mangachena and Pickering (2021) who found that of the 229 mammal species in South Africa, tourists most wanted to sight elephant (48%), black and white rhino (36%), lion (36%), cheetah (35%), and giraffe (34%).

Ultimately, it seems that species is the primary motivator for uploading a sighting, as collections of images and focus groups responses all showed/stated that a low-quality image of these iconic species would be included in a 10-image collection at the expense of a higher quality image of a "less interesting" species (Mangachena and Pickering, 2021). Focus group respondents stated that (and image collections recorded proved) they would upload a high-quality image at the start of a collection, with less emphasis on the species, but that the subsequent inclusion of content would be determined heavily by the subject matter; be that a highly anticipated species or a particular experience. It seems that this philosophy did not extend as far as to encourage the posting of multiple images of the same sighting within a collection, and that rather than multiposting a sighting of a highly anticipated species, users would post content of another sighting or experience.

5.9.1 A note on Rhinoceros and Pangolin

Although southern white rhinoceros, black rhinoceros, and pangolin species are extant in both parks, they were not discussed here in any depth to protect them against the high levels of poaching and trading they are victim to. To appreciate how online spaces can be used to study these species and better their conservation, and conversely how their presence online might endanger them in the current situation, it is important to briefly, if intentionally vaguely, discuss their representation.

Southern white rhinoceros were the 12th most posted species and black rhinoceros the 21st most posted, indicating a high interest in rhinoceros on social media. LSApp prohibits the upload of rhinoceros sightings on their platform. Subsequently, removing LSApp sightings and reposts as a consequence thereof, white rhinoceros become the 8th most posted species. This indicates that white rhinoceros are both popular on online platforms and easily sighted within parks. Black rhinoceros sightings are significantly less prevalent, accounting for about 9% as many sightings as white rhinoceros. Pangolin were also restricted on LSApp, with sightings having their location data removed. Pangolin were only represented on LSApp (and subsequently reposted to Twitter and WhatsApp). Behaviours among these sightings were distributed similarly to top species, meaning that sightings are still dictated by increased chances of and duration of exposure as dictated by these behaviours, and they were presented in similar distributions of content type (apart from Pangolin where all sightings were image, though there were only a handful to select).

Rhinoceros conservation can cost inordinate amounts of time, money, and resources (Rowan, 2023). For large reserves, monitoring the locations, behaviours, and health of their rhinoceros can be logistically impossible, and the majority of their protection relies on the general maintenance and upkeep of security and fences. Survey methods that estimate populations of all animals within a park are broadly expensive (as discussed in section 1.2) and as such are often conducted on an annual basis, if not even more infrequently. The application of these methodologies could result in constant, inexpensive monitoring of these endangered populations without relying on geotagging the animal or recapturing them.

Subsequently, parks could come to an arrangement with the Latest Sightings team to allow rhinoceros sightings to be reported on LSApp but, rather than posting them to a timeline, any and all sighting information could be securely reported to the parks' rangers to aid their efforts, as they did with pangolin locations. Unfortunately, due to the value of rhinoceros horn, even possessing information about their whereabouts could endanger individuals. The level of organisation within the criminal groups that poach rhinoceros means that acting as a conduit for this information could endanger members of the Latest Sightings team (Van Uhm, 2012; Montesh, 2013) and the rangers seeking to protect them.

If this compromise is considered too dangerous to those protecting the rhinoceros, or alongside this approach, it might also be decided that having any amount of rhinoceros sighting posts online is a significant danger to the species. As such, an informed and considered campaign should be conducted on the platforms and/or at the parks to reduce the number of rhinoceros sightings, or at least ensure that the content provided does its best to anonymise the sighting by removing location data and ensuring no landmarks in the posts or any other posts on the account inadvertently reveal the location of these species.

5.10 Data Types: Individual Count

Species count data ultimately proved insufficient to complete species abundance estimates (Madon, *et al*., 2011). Although there was nothing inaccurate about the species count data themselves, a sufficient quantification of sighting recapture was not employed and as such the overrepresentation of popular species occurred. To combat this, a separate study would need to quantify the rates at which individuals of each species are recaptured by identifying groups of individuals and recording all sightings and posts referencing that group. Doing so would allow for the definition of species quantity estimates by the rates of recapture and reposting, allowing for accurate population estimates to be obtained. A study of this specificity, depth, and magnitude would likely also uncover other sources of inaccuracy within species count data, with mitigation of these errors also aiding in reduction in sources of error for the methodology employed here.

5.11 Data Types: Additional Species

The presence of additional species in posts can be used to monitor population densities, how species react to features such as water or salt licks, symbiotic interactions, and myriad other ecological phenomena (Eksteen and Bornmann, 1990). The additional species data collected here ultimately proved that there were biases towards the primary species a sighting was made for. The contingencies described in section 3.9 were used to evaluate the biases of the content creator without imposing any personal biases through the data collection process, similar to attempts to standardise camera trap setup procedures to reduce bias (Hofmeester, *et al*., 2019; Mills, *et al*., 2019). Ultimately the most popular species here are also the subject of the most bias, in turn causing greater contribution to their overrepresentation.

5.12 Data Types: Behaviour

Behaviour is perhaps just as influential in the decisions OCs make to record and report a sighting as species. The behaviour an animal expresses can serve to increase or decrease its visibility or the duration of its exposure to an OC, increase or decrease the desire of an OC to record it and the content through which they may achieve that, and even impacts the propagation of a sighting to platforms.

The content preferences and communities on Instagram seems to have impacted the animal behaviours being captured by Instagram users and subsequently presented on the platform. On non-Instagram platforms, feeding behaviour accounts for 12% of leopard behaviours, 7% of lions, and a staggering 24% of elephant post behaviours. On Instagram, however, the behaviour accounts for just 3% of leopard and 4% of lion behaviours, but still 24% of elephant. Giraffe, zebra, and antelope also share much closer representations of feeding behaviour on Instagram compared to other platforms. This suggests a level of censorship among carnivores on the platform. Carnivore kills and feeding are often considered to be gory or “raw” wildlife sighting experiences, and overall do not seem conducive to Instagram’s content posting guidelines and the content which its algorithms promote.

Facebook, on the other hand, has a powerful emphasis on such content. Likely due to the prevalence of groups which allow users to more easily self-moderate the types of content they present, and which allow for content to only be shared among users who are unlikely to take issue with it, carnivores are increasingly prevalent on the site. Facebook contained the most video and long-form video sightings of any platform (accounting for 92.71% of long-form video sightings), many of which were used to present (carnivore) feeding behaviours (more prevalent on Facebook than all platforms apart from Flickr) or intraspecies interactions (which were more prevalent on Facebook than any other platform). Because of this, videos on Facebook can be described in one of three ways: (1) “brutal nature” videos where predators hunt, make a kill, predate on an already dying animal, target the young of another species, or eat an animal that is still struggling; (2) “the underdog” escapes or the prey turns on the predator to scare them off or greatly injure them; or (3) “family” interactions between animals of the same species; typically lion and elephant.

Due to the frequency of social media users to personify and empathise with animals this content is particularly polarising and garners hundreds of comments and reactions per post, seemingly more so than any other content and behaviour pairing. Facebook is the only place where closed groups protect this form of content, which is likely why these behaviours are more prevalent here than anywhere else. The Latest Sightings team, for example, do not post this sort of content themselves and do not advertise sightings of this nature any more so than other behaviours when reposting from LSApp. Instead, they seem to use this polarising content to entice high engagement on Facebook posts (Garimella, *et al*., 2018) and redirect users to their other pages and platforms where they can gain donations and users.

Furthermore, species locomotion increased their chances of intersecting with roads and tracks, and subsequently vehicles and OCs. Roads can serve as wildlife passages, particularly for mammals (Little, *et al*., 2002), have potentially nutrient rich and palatable grazing on their verges (Safford and Harrison, 2001), and as such can often attract prey animals (Fahrig and Rytwinski, 2009) and subsequently predators. All of these factors would serve to increase the chance of sightings being made in the parks due to the limitations posed on visitors to remain in their vehicles and for those vehicles to remain on the roads. Prey animals such as antelope or zebra could act as an indicator of a predator’s presence if they were exhibiting alert or skittish behaviour as tourists approached, which would likely lead to an increase in vigilance and/or encourage visitors to remain in that area, on the lookout for a desired carnivore sighting. If and when an instance of predation occurred, this would likely become the subject matter of any reported sighting, where the carnivorous species ultimately supplanted the herbivorous species which was initially sighted.

5.13 Data Types: Location

Location data as provided in the text descriptions of posts and attached to LSApp tings is too contextual, circumstantial, and varied to be applied to any subsequent study. The tings provided by LSApp on their ting map, however, provide current and relevant information using a more reliable source of location data from the location services of the devices using the app, which are typically accurate to a matter of metres. It proved elementary to correlate a ting on a map to the fully detailed sighting regarding it, which made it possible to track the movements of individuals who (a) possessed distinguishing features such as injuries or markings and (b) were sighted frequently. Location data in this format could be used to track the movements of individuals, which could prove particularly useful due to the focus of sightings towards the perimeter of PNP.

Perimeter fences are the most concentrated areas for poaching incidents. Increased visibility to perpetrators, falling within the range of firearms, and reduced exposure of perpetrators reducing the risk of detection leads to this increased number (Cambron, *et al*., 2015; Kamminga, *et al*., 2018). There were a large number of high-concentration and individual sightings around the perimeters of PNP. These correlated, mostly, with the presence of roads and pockets of moderate to strong signal, but still suggest populations of animals that congregate around these fences. Tracking the movements of animals could therefore not only provide behavioural insights, but serve to inform and improve conservation efforts.

5.14 Data Types: Search Terms

Two formats of search terms were the most effective on different platforms. The default search functions of Twitter and Instagram, which support both formal and informal #hastagging systems, returned a greater number of relevant results when these #hastags were employed alongside the use of single key words or initialisations of the parks. Instagram's location tagging format also means that sightings could be searched for by utilising the platform's location lists. Otherwise, "#KNP/PNP sighting" yielded the highest concentration of relevant content on these platforms. On these platforms, search terms such as "Kruger/Pilanesberg National Park" proved too long-winded for the search functions and introduced large numbers of irrelevant content.

LSApp and WhatsApp did not require search terms due to the enforced relevance of content upon them, but the search in chat functions WhatsApp offers were useful for returning to backdated posts and attempting to verify suspicions of reposts or multiposts.

Search terms on Facebook seemed to be of lesser importance due to the membership of accounts within groups (the joining of which conformed to the limitations described in section 3.2.1 and 3.12.1 which were implemented to protect the privacy and wellbeing of groups members). After a search term had been input, filters could be applied to the results which include searching from all account types, just pages, or just groups. Ordering by recent and filtering to pages and groups proved to remove most of the irrelevant content, and a second search conducted including personal accounts but excluding groups and pages was subsequently completed to ensure no posts were missed (though very few were discovered in this second sweep).

Flickr's search functions are similarly generous. Filtering by taken on date to ensure that post lag was not so great that post content was no longer relevant to the study did not remove a noticeable amount of content when compared to searches not implementing this technique. Rather, organising posts by taken on date served to simplify the data input process due to the particular layout of the spreadsheet used to record observations, but this was not essential.

The inclusion of the word "sightings" seemed essential to collect most of the posts observed in this study. Very few other search terms, including those referring to safari and wildlife, returned new or relevant results. No non-English search terms were studied in this project, and the inclusion of other languages could have returned new and unique results, as opposed to attempting alternative English phrases.

5.15 Signal Coverage and Post Location

The game drive and signal coverage testing conducted in PNP provided valuable insights into the ability of users to create content regarding a sighting. Figure 21 presents the findings of these tests, and their correlation with sighting density based on LSApp tings. Although the majority of strong signal pockets were retrieved within the large, concentrated bubble of sightings, so too did the majority of pockets of no signal. This implies that the correlation may not lie between sightings and signal coverage, but rather between sighting and effort.

Being that the majority of PNP's visitors are day guests, there are a finite number of hours OCs have the potential to be exposed to a sighting. Sighting time data implies that sightings are more likely to occur in the mornings, which somewhat correlates with the ethology of most species, but does not present the full scope of this trend. Rather than actually determining the likelihood of a sighting being made, it is here proposed that the spatial distribution, count, and temporal distribution of sightings is instead determined simply by the interest and motivation levels of the visitor.

The distribution of sightings in PNP corresponds with the locations of main roads. These main roads in the park are linked to three of the park's five gates, and the main three that would be used by day tourists. The gate in the northwest, for example, primarily serves the Black Rhino Lodge, and as such is less suitable for general admission. Subsequently, the majority of visitors to PNP will arrive via the gates indicated to the north, northeast, and south of figure 21, and immediately find themselves on the main roads. These roads are the most convenient modes of travel for many visitors as they traverse large swathes of the park and are maintained to a higher standard than many of the side tracks. Therefore, accessibility is also increased. These roads also have relatively more consistent topography than many of the side roads, increasing visibility, intersect with much of the park's infrastructure in the forms of hides and the Pilanesberg centre, and cover a range of the park's environments from veld to acacia shrub to large bodies of water (dams). There is also likely a positive feedback loop occurring wherein increased traversal of these roads leads to increased sightings, reporting of these sightings leads to increased traversal, and so on.

Subsequently, the distribution of sightings along these main roads seems logical. For visitors who have visited PNP before, or whose vehicles are more capable, the prospect of venturing from these paths leads to the traversal of the lesser tracks. These tracks take longer to cover the same distances, due to the uneven surfaces, and cover larger distances than the main roads. Subsequently, by the time a visitor is at the furthest section of a lesser road, another visitor may have covered a greater distance on the main roads. Furthermore, the main roads would have to be traversed to access these side tracks anyway. This correlates to the distribution of sightings which, although present over PNP's side roads, seem to spread out from the main roads of PNP.

It therefore seems that the likelihood of a sighting being recorded and/or reported may not correspond to location, but rather to effort which itself corresponds to relative location within the park (distance from gates, main roads, and amenities) (Ball, *et al*., 2023). The time a sighting is made can subsequently give an indication of the level of effort expended for that sighting to be made. Increased posts in the morning would correspond with visitors traversing the main roads of PNP, through the high-density bubble. After visitors have spent longer in the park, they venture out onto the smaller roads, decreasing concentration of visitors and subsequently sightings. All of this occurs over a matter of hours, wherein visitors will be desensitised to certain sightings, demoralised by their inability to upload them, or more concerned with the wellbeing of their vehicle or planning a route that allows them to exit the park before gates close and they incur a fine.

The sporadic sightings present to the left of the area of highly concentrated sightings are a series of points towards Lenong View where visitors could stop to rest (or might decide to turn around) and would subsequently have more time to record or report a sighting. The relative niche of the biome also contributes to new species detection at these locations, further increasing motivation to upload a sighting. Similarly, hides and dams, which allow for visitors to rest, check their devices, and attempt to upload sightings, may also allow for sightings of different species, such as those which favour water (hippos or some reptilian and amphibious species).

Although these observations are somewhat speculative, they are supported by the distribution of posts in PNP, where location corresponds to sightings, and would also correspond to effort spent; in this case time spent in the park. Where users are preoccupied with navigating lower-quality roads, have been desensitised by multiple viewings of the same species, or demoralised by lack of access to network connectivity, motivation to create a sighting will decrease, which correlates with the decrease in sighting reporting presented in this study's main datasets.

A review of visitor experiences, accompanied by surveys similar to those conducted here but with an increased focus on the experience of a tourist in PNP over time would serve to verify these concepts. These would also allow for a wider array of insights into network connectivity as qualitative assessments of a more varied collection of network providers and devices could be obtained relatively easily. A study into the correlations between effort and sighting should be conducted on both behaviour and species to verify the proposal of this project that increased exposure is a contributing factor to a sighting being made and that as effort (time spent in park) increases, so too do expressions of favouritism by visitors as they become more selective about what they report.

5.16 Collecting Sighting Data from Social Media Platforms

The particularities, features, communities, restrictions, guidelines, and general uses of a platform influence all content that is uploaded to them, including those of wildlife sightings. Many of those were expressed in the data through content type, account type, and subject matter. These characteristics differ platform to platform and are the root cause of the variation of data recorded upon them.

5.16.1 Twitter

Despite supporting many of the features, content types, and community structures that wildlife sightings would thrive on, Twitter produced less unique sightings than Instagram and Facebook, and only twice as many sightings as Flickr despite boasting ten times as many concurrent monthly users. This demonstrates an under-utilisation of Twitter by wildlife tourists in KNP and PNP, which is likely caused by several factors.

The recent rebranding of Twitter to X has led to a mass exodus of the platform. Since Elon Musk's purchase of the platform in October of 2022, Chang, *et al*. (2023) estimate that up to 50% of environmentally conscious users have abandoned the platform in protest of the unsustainable business practices Musk employs in his development of technologies through Tesla and SpaceX (Pierson, 2021). This boycotting likely includes a large number of safari guests, and such experiences would likely attract a concentration of environmentally conscious individuals than a typical holiday.

The lack of wildlife sightings on Twitter is still odd, however, as average ages of Twitter users correspond with the average age of PNP visitors at 25-34 years of age and 34 years of age respectively (Sloan, *et al*., 2015; Nikodinoska, *et al*., 2014). It seems therefore that it is more likely the features of Twitter that causes this under-utilisation. The high demands for device connectivity discussed in section 5.15, the type of content generally held on the app, and the tendency to be used for informative and professional purposes opposed to socialising with friends and family. These lead to a lack of mammal sighting information to be collected from passive studies such as this, but could potentially serve to improve response to studies relying on public outreach due to the high levels of engagement and propagation of information upon the platform.

5.16.2 The Latest Sightings Mobile Application

LSApp is purpose built to report wildlife sightings, and it serves this purpose well. LSApp receives 1,000-2,000 daily users, which equates to about 20,000 unique active users monthly (Latest Sightings, 2023, b). Due to the nature of the application, the majority of these users are only engaging with the application while on their holidays and as such are likely to be inactive for large periods of time, if they ever return to LSApp at all. This userbase is considerably smaller than that of any other social media platform, where Flickr has 60,000,000 monthly users; Twitter 556,000,000; Instagram 2,000,000,000; WhatsApp 2,000,000,000; and Facebook 2,985,000,000 (Statista, 2023). This truly puts into perspective how effective LSApp is and how dedicated its users are. LSApp routinely contributed more daily sightings than Instagram and Facebook combined, and generated more posts every day than all platforms combined due to the frequency of reposts from LSApp to Twitter and WhatsApp.

Ultimately, this leads to LSApp being the greatest source of mammal data of any platform studied here. LSApp produced the most extensive species inventory and provided the most reliable sighting day and sighting time data. The ting map of LSApp was also the only reliable way to spatiotemporally reconstruct sighting locations, and the LSApp community broadly recorded sightings of many obscure or elusive species that the platform's users seemed to actively seek out (Blenkinsop, 2018). On a platform where lion, elephant, and leopard sightings are so common, it seems that the presentation of new and unique species is increasingly appealing among the platform's users (Blenkinsop, 2018).

That said, LSApp still exhibits prevalent biases towards many species. The majority of its users will be relatively casual tourists, who will be hoping for the same calibre of experiences described in section 5.6. Subsequently, LSApp may be the best platform to capture large amounts of data easily, but it still exhibits many of the same issues for data quality that other platforms present.

5.16.3 Instagram

Instagram is predominantly used by young adults and is the preferred platform of this age group (Pittman and Reich, 2016). Content on Instagram is incredibly varied, with up to 15% of daily posts concerning activities such as daily errands, attending events, travelling and holidays (Hu, *et al*., 2014). It is unclear as to whether this is causal or consequential of the interpersonal influencing that occurs on the platform. Regardless, it is apparent that Instagram is the platform used for young people to document their experiences, including wildlife sightings.

The desire of Instagram users to depict their lives as glamorous and desirable (peacocking) is not conducive to sharing subject matters that wildlife sightings often contain. Beyond the limitations of the platform to images and videos, Instagram users seem to generally steer clear of uploading sightings revolving around carnivore feeding, hunting, or other interactions that may involve an animal being harmed (Campbell, 2019). The reduced representation of these events was replaced on Instagram with portrait posts; though these portraits were often of lions, leopards, and spotted hyenas; animals that exhibit the very behaviours that are apparently censored on the platform. It seems therefore, that predators are not a concern to Instagram users, but predation is. This level of censorship is potentially detrimental to the quality of the dataset as it impacts the integrity of behaviour data collected, evident by Instagram's moderate deviation from the reported norms of this type of data.

5.16.4 Facebook

Facebook groups served to concentrate data, and also act as a means of moderating the integrity of sightings posted to it. Groups are spaces where users and content are isolated from the majority of the platform's userbase, and where all members understand the intended content of the space. Subsequently, there are instances of content in Facebook groups that might not be posted to other platforms or non-group spaces as they would likely be reported by distressed users and removed. Facebook was the source of the majority of long-form videos pertaining to animal hunting and predation, likely for these very reasons.

Accessing group content on Facebook ensures relevance of the data but can also serve to assure the observer that there are less instances of censorship taking place; potentially leading to a more accurate representation of animal behaviours on the platform. Unfortunately, Facebook expresses the same bias to species as other platforms, somewhat negating this level of integrity, and will have similar signal demands to Twitter meaning that uploads likely experience post lag that cannot be quantified due to an overall absence of sighting day data on the platform. Furthermore, Facebook posts did not contain any location data; though if they did, based on the text descriptions on posts, it would likely conform to the descriptive format expressed on other platforms and subsequently have limited application in conservation studies.

5.16.5 Flickr

Flickr shared a level of dedication among its users with LSApp, although their dedication was towards the hobby and art form of photography (Tankanen, *et al*., 2017). Mammal sighting posts on Flickr were of noticeably higher quality than any other platform. Post lag was also significantly greater here, and it seems that these two factors are connected. Due to the high quality of Flickr images, uploads are likely completed upon an OC's return to their home where they can transfer images from their camera to personal device easier and have a familiar and reliable internet connection. During this process, users likely curate the content that they post. This will include selecting the best photos of a sighting to present and could also include editing which would further extend the process and increase post lag.

That said, the calibre of content on Flickr, not only its composition but also the quality of data, was second only to LSApp. Sightings used the higher image quality and tagging system to annotate additional background details and highlight additional species sightings that may have been missed or possesses high uncertainty on other platforms due to the pixelation of images as they were zoomed in on. The taken-on date served as the only instances of non-LSApp sighting day data, and there were frequently attempts to include location data on Flickr. Additionally, the posts observed were mostly from the collections of just a few creators, uploading dozens of images each, implying that Flickr has great potential in terms of the quantity of data it can present though, unlike Instagram and Facebook where the algorithms present less posts the further one backdates, Flickr does not seem to disregard older posts, meaning that backdating is more accessible, but that the posts collected here represent the full discoverable amount of Flickr's content for the study period.

5.16.6 WhatsApp

As a messaging service, WhatsApp provides fewer social features than other platforms in return for greater moderation and privacy. The only relevant WhatsApp groups found were those of the Latest Sightings team, which subsequently led to a lack of new content on the platform. WhatsApp still proved useful in the data collection process, however, as the daily summaries posted by the Latest

Sightings team included screenshots of the ting maps that could otherwise not be backdated and presented updates on activities in the parks such as the organisation's affiliation with the pangolin research project.

The chats also provided frequent updates for weather, road closures, and other incidents which could serve to inform research groups conducting their research in person. Similar information was tinged on LSApp, also, but due to the layout of the ting timeline was easily buried.

Although WhatsApp best served to augment the collection of LSApp data and to verify suspicions of reposts and multiposts from LSApp, the collection of data on the platform and the increased ability to backdate in the group chats presents great potential to be useful in many applications beyond what this methodology employed.

5.17 Review of the Manual Approach

The process of collecting extensive social media data demanded significant time, labour, and effort. Over several weeks of dedicated research sessions (lasting 8+hours each), efficient systems of data collection and recording were established. These initial efforts involved tasks such as identifying relevant Facebook groups, devising the platform-specific search terms, and familiarizing oneself with the nuances of each platform. Furthermore, defining specific parameters, such as the hierarchy for determining primary species and the definitions of observed behaviours, required significant due diligence. Upon reflection, these tasks likely consumed around 72-hours of work (these efforts were spread across approximately three weeks including rest days and days where no progress could be made due to other commitments).

Once these foundational frameworks were established, data collection commenced. Techniques were devised within the data recording documents to facilitate easier and more efficient data collection. Firstly, the categories that were recorded were ordered in a way that meant data could be input in the order it would be collected from a post. For example, the left-most column was Date Accessed, then Platform, then Post Date. These factors were the least likely to change from one post to another and subsequently could be left blank (unless they did change) and infilled with the relevant data after several posts had been recorded. Similarly, Sighting Date and Sighting Time were placed towards the left of the document, as these were often not obtainable from the context provided.

All of the parameters which were always recorded were grouped in the middle of this document, with Account Type, Content Type, and repost preceding species-specific data; Species, Species Quantity, Behaviour/Portrait, Location, Multipost. These were the essential datapoints that a post had to contain to be recorded for this study, and were placed in the order they would become most apparent from the post. Park and Search Term were next as, as with Date Accessed, etc, these variables were controlled by the researcher and unlikely to change. Cells that recorded additional species were placed to the right of these parameters as they were the least frequently required.

This framework of data input meant that a workflow could be quickly developed, and that a relevant post could be processed methodically and efficiently. This data collection process was again augmented by the computer peripherals that the researcher employed. Many computer mice (in this case the Corsair Scimitar) have additional buttons that can be programmed in dedicated software. In this instance, a 12-digit number pad on the side of the mouse left a total of 17-programmable buttons. Some of these had alternate keystrokes prescribed to them to streamline data input. “7” was assigned to “Tab” to move between input cells quickly. Similarly, “11” was mapped to “/” and “12” to “:” so that imputing dates and times was quicker. Most significantly, “5” was mapped to instantly resolve 9 keystrokes; Down Arrow and then eight instances of the Left Arrow, which would take the cell selection from the Multipost column to the Account Type column of the next row, which was the most common starting point for data input of the next post.

This methodical approach, applied to over 15,000 posts and 220,000 datapoints (most posts contained 14 parameters of data, while some contained up to 26), resulted in substantial time savings equivalent to hours of labour. With practice, a day’s data collection for all six platforms across both parks could be completed in two hours. This productivity was occasionally hindered by a need to backdate. Although possible and no more complicated than punctual data collection, data collection sessions where multiple days’ posts were recorded became fatiguing and took exponentially longer. The longest instance of backdating was for a period of 16 days (covering the time the researcher was on fieldwork) which took upwards of 48 hours (8*6-hour days) to compensate for, including the recording of new posts within that time.

Based on the worktimes described above, data collection for this study likely took upwards of 212 hours. The reference document which recorded the behaviour definitions and alphanumerical coding had a total edit/open time of 9,780 minutes (163 hours), though was not always opened at the start of data collection, and had multiple previous iterations.

In this time, this manual approach compiled a 55-species mammal inventory across both parks, but a similarly impressive 44-mammal species inventory could be completed in 29 days (figure 6). The researcher also gained valuable insights into the communities of each individual platform, their interests and preferences (illustrated in section 5.16.1 through 5.16.6), and the ways which they communicate. This level of familiarity with the content being analysed not only assisted in interpreting the data, but also formed an appreciation for these online communities which still remain relevant to the successful application of the data to conservation efforts (Danielsen, *et* al., 2007; Hausmann, *et al*., 2017; Jarić, *et al*., 2021). Although rapidly advancing Artificial Intelligence (AI) technologies may soon be able to quantify relationships between species representation and people and platform, it may not be able to communicate the qualitative aspects of these data to a researcher to the extent where they can write a well-informed and insightful paper or where a conservationist can identify multi-faceted threats to a system. It is here that the manual approach would allow for a human-element of insight and practical reasoning to better interpret these data.

Similarly, the identification of a sighting as a multipost required a level of thought beyond current AI capabilities. Here, content was not matched, but rather the content of multiple different posts was assessed, and a considered judgement was made as to whether, based on the balance of probabilities, each post was independent or a separate post of the same sighting. This is where the presence of a person can add value to data collection at a level above a computer.

5.18 Implications for Wildlife Conservation

At present, discussions around potential applications of digitally sourced wildlife data are multifaceted (Norouzzadeh, *et al*., 2018). While the transition of the field into a “big data” science appears inevitable and advantageous, this study contributes a manual retrieval methodology to this ongoing conversation. Although automated collection methods relying on APIs and deep learning are more efficient, this manual approach significantly reduces the threshold required to access wildlife-relevant data from social media posts. By eliminating the need for advanced technical skills, this methodology opens avenues for ecologists of many backgrounds and levels (including students and those who may not have opportunities to collaborate with academics with strong technical backgrounds). Furthermore, this manual approach is a cost-effective alternative to existing ecological survey techniques for researchers with smaller-scope studies or limited resources.

Additionally, this study acknowledges the voluntary effort of content creators in generating social media content, highlighting the importance of respecting their contributions in-line with ongoing discussions about how “the user” is observed by scientists in online spaces (Budhathoki, *et al*., 2008; Leadbetter, 2008; Reed, 2020). Moreover, the study’s focus on user habits and platform preferences appreciates the importance of qualitative factors of the data (Di Minin, *et al*., 2013), similar to conservation culturomics. By integrating ethical considerations and cultural insights into this manual approach, the methodology hopes to offer a holistic and accessible approach to wildlife data collection in the information age.

Section 6: Concluding Remarks

6.1 Conclusion

This methodology has presented great potential for data collection to achieve wildlife conservation. Mammal sightings data can be manually obtained, in a vast quantity and to a reasonable quality, from publicly accessible social media posts. Retrieving these posts can be done in a way that is accessible, relatively effort effective, and while respecting the expected privacy of social media users. All of this was achieved and can be replicated utilising the default search functions of a social media platform, without any necessity for investment in APIs or other external software or the development of associated technological skills.

A study which identified individuals across social media posts and could as such quantify rates of recapture across species would serve to improve the applicability of this dataset greatly; allowing for the vast quantity of content obtained to be applied to population estimates and analyses. These numbers are essential to conservation efforts and reserve management, but the identification and tracking of an individual animal can also be employed to track their wellbeing, ensure the safety of at-risk individuals, or even follow pregnant or injured animals where intervention may be required.

The habits expressed by ecologically minded individuals in this study not only correspond with the findings of the dataset but also with preexisting notions of social media usage from past literature. A combination of personal preference towards species and their behaviours, as well as access to internet connection and what platform an OC uses dictates how an event is captured and to what extent, and also how and where these sightings are propagated as additional posts. Animal abundance plays a less significant role in the creating of a sighting online as the most abundant species are routinely under-represented due to an apparent desensitisation of OCs to these sightings, whereas elusive and low-population species sightings seem to always encourage a post to be made. Species behaviours are not only sought after as the content of a potential post, but can also be causal to a post's existence as behaviours that encourage the convergence of an animal with roads increase the likelihood and duration of exposure of an animal to a content creator.

Ultimately, the greatest application of this methodology was proven to be the creation of a species inventory for a region or park, where high-detail location data would not be required, and where potential bias of species, censorship of behaviours, and recapture of individuals would be less relevant. These data can be used to inform conservation projects in many ways such as scouting to confirm a species is extant before a project regarding them is committed to, monitoring visitor interests and subsequently inspire potential campaigns for fundraisers, and could be adapted to monitor individual animals or groups to ensure their safety or observe their behaviours.

This methodology is widely applicable, and has the ability to be adopted by researchers at many levels where automated collection may be exclusive. Additionally, the ability of a researcher to compensate for the context around digital media, particularly the existence of multiposts and reposts which current machine-learning capabilities cannot compensate for. Although there are many benefits to wildlife conservation adopting automated data collection from digital sources to become a “big data” science which this study does not rebuke, there is also certainly still a place and niche for the manual retrieval of data from such sources. Integration of this approach into the iEcology belief system can not only collect accurate and reliable data, but can also serve to grant researchers a level of insight and appreciation of the data that they may otherwise miss.

References

Ahmed, W., Bath, P. and Demartini, G. (2017), 'Using Twitter as a Data Source: An Overview of Ehtical, Legal, and Methodological Challenges', *The Ethics of Online Research. Advances in Research Ethics and Integrity* (2), UK:Emerald, pp.79-107

Alexander, L. (2016), 'Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond', *Weather and Climate Extremes*, 11, pp. 4-16

Anderson, S., *et al*. (2021), 'Trends in ecology and conservation over eight decades', *Frontiers in Ecology and the Environment*, 19(5), pp. 274-282

Arts, K., van der Wal, R. and Adams, W. (2015), 'Digital technology and the conservation of nature', *Ambio*, 44(4), pp. 661-673

Bakshy, E., Hofman, J., Mason, W. and Watts, D. (2011), 'Identifying influencers on Twitter', *Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, 2

Balaji, T., Rao Annavarapu, C. and Bablani, A. (2021), 'Machine learning algorithms for social media analysis: A survey', *Computer Science Review*, 40, pp. 10-19

Ball, I., Marneweck, D., Elliot, N., Gopalaswamy, A., Fritz, H. and Venter, J. (2023), 'Considerations on effort, precision and accuracy for long-term monitoring of African lions (*Pantera leo*), when using Bayesian spatial explicit catpure-recapture models, in fenced protected areas', *Ecology and Evolution*, 13(7)

Barrow, E. and Murphee, M. (2001), 'Community Conservation: from concept to practice', *African wildlife and livelihoods: the promise and performance of community conservation*, pp. 24-37

Baselga, A. (2013), 'Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs abundance gradients', *Methods in Ecology and Evolution*, 4(6), pp. 552-557

Bausch, P. and Bumgardner, J., 2006. *Flickr Hacks: Tips & Tools for Sharing Photos Online*. " O'Reilly Media, Inc.".

Beeton, N., *et al*. (2015), 'Using the spatial population abundance dynamics engine for conservation management', *Methods in Ecology an Evolution*, 6(12), pp. 1407-1416

Bercovitch, F. and Berry, P. (2013), 'Herd composition, kinship and fission-fusion social dynamics among wild giraffe', *African Journal of Ecology*, 51(2), pp. 206-216

Blenkinsop, S. (2018), 'The effects of wildlife applications on the visitors' authentic nature experience in the Kruger National Park', *ISTA 2018 Conference Proceedings*, pp. 15-22

Bottrill, M., *et al*. (2009), 'Finite conservation funds mean triage is unavoidable', *Trends in Ecology and Evolution*, 24(4), pp. 183-184

Boyle, S., Kennedy, C., Torres, J., Colman, K., Pérez-Estigarribia, P. and de la Sancha, N. (2014), 'High-resolution satellite imagery is an important yet underutilized resource in conservation biology', *PLoS One*, 9, p. e86908Braje, T. (2016), 'Evaluating the Anthropocene: is there something useful about a geological epoch of humans?', *Antiquity*, 90(350), pp. 504-512, doi: 10.15184/aqy.2016.3

Britannica (2023), 'Kruger National Park', *Geography and Travel*, Available at: https://www.britannica.com/place/South-Africa/Relief (Accessed: Aug 2023)

British Psychological Society (2021), *Ethics guidelines for internet-mediated research,* UK:British Psychological Society

Brooks, T., et al. (2002), 'Habitat Loss and Extinction in the Hotspots of Biodiversity', *Conservation Biology*, 16(4), pp.909-923

Brown, R. (2002), 'Innovations for conservation and development', *Geographical Journal*, 168, pp. 6-17

Browning, E., Gibb, R., Glover-Kapfer, P. and Jones, K. (2017), 'Passive Acoustic Monitoring in Ecology and Conservation', *Conservation Technology*, Woking, UK: WWF-UK

Budhathoki, N., Bruce, C. and Nedovic-Budic, Z. (2008), 'Reconceptualizing the role of the user of spatial data infrastructure', *GeoJournal*, 72(3/4), pp.149-160

Bush, E., Baker, S. and Macdonald, D. (2014), 'Global trade in exotic pets', *Conservation Biology*, 28(3), pp. 663-676

Butchart, S., et al. (2019), 'Chapter 3: Assessing progress towards meeting major international objectives related to nature and nature's contributions to people', *The Global Assessment Report on Biodiversity and Ecosystem Services*, DOI: 10.5281/ZENODO.6417320

Caine, J. (2007), 'Can Food Supplements affect Ulcerative Skin Disease in Black Rhinoceros (Diceros bicornis)?: An Experimental and Chromatographic Study', *8th Annual Symposium on Zoo Research*, pp.32-39

Campbell, M. (2019), 'Conservation psychology, social media and the conservation of large carnivores', *Advances in Conservation Psychology*, Nova Science Publishers Inc: New York, USA. Pg. 1-47

Cambron, M.E., Brode, C., Butler, P. and Olszewkski, G., 2015, April. Poacher detection at fence crossing. In SoutheastCon 2015 (pp. 1-2). IEEE.

Caro, T., Darwin, J., Forrester, T., Ledoux-Bloom, C., and Wells, C. (2011), 'Conservation in the Anthropocene', *Conservation Biology*, 26(1), pp. 185-188

Caughley, G. (1977), 'Sampling in aerial survey', *The Journal of Wildlife Management*, pp. 605-616

Cernev, T. and Fenner, R. (2020), 'The Importance of achieving foundational Sustainable Development Goals in reducing global risk', *Futures*, 115, pp.1024-1092

Chang, C., Deshmukh, N., Armsworth, P. and Masuda, Y. (2023), 'Environmental Users Abandon Twitter after Musk takeover', *Trends in Ecology and Evolution*, 10(10), pp.19-21

Chapman, F. (1927), 'Who treads our trails?', *Natural Geographic Magazine*, 52, pp.330-345

Chatterjee, A.B., Sankar, K., Jhala, Y.V. and Qureshi, Q., 2023. Spatio-temporal patterns of co-occurrence of tigers and leopards within a protected area in central India. *Web Ecology*, *23*(1), pp.17-34.

Chua, A., Servillo, L., Marcheggiani, E. and Moere, A. (2016), 'Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy', *Tourism Management*, 57, pp. 295-310

Clancy, T., Pople, A. and Gibson, L. (1997), 'Comparison of helicopter line transects with walked line transects for estimating densities of kangaroos', *Wildlife Research*, 24(4), pp. 397-409

Costa-Pereira, R., Moll, R., Jesmer, B. and Jetz, W. (2022), 'Animal tracking moves community ecology: Opportunities and challenges', *Journal of Animal Ecology*, 91(7), pp. 1334-1344

Creevey, D., Coughlan, J., and O'Conner, C. (2019), 'Social media and luxury: A systemic literature review', *International Journal of Management Reviews*, 24, pp.99-129, DOI: 10.1111/ijmr.12271

Curtin, S. (2010), 'Managing the wildlife tourism experience: The importance of tour leaders', *International Journal on Tourism Research*, 12(3), pp. 219-236

Danielsen, F., et al. (2007), 'Increasing conservation management action by involving local people in natural resource monitoring', *AMBIO: A Journal of the Human Environment*, 36 (7), pp. 556-570

Deng, C., Daley, T. and Smith, A. (2015), 'Applications of species accumulation curves in large-scale biological data analysis', *Quantitative Biology*, 3(3), pp. 135-144

Davis, C. (1982), *CRC Handbook of Census Methods for Terrestrial Vertebrates*. Boca Raton, Florida:CRC Press Inc.

Denninger Snyder, K. and Rentsch, D. (2020), 'Rethinking assessment of success of mitigation strategies for elephant-induced crop damage', *Conservation Biology*, 34, pp. 829-842

Di Minin, E., Fraser, I., Slotow, r. and MacMillan, D. (2013), 'Understanding heterogenous preference of tourists for big game species: implications for conservation and management', *Animal Conservation*, 16, pp. 1-10

Di Minin, E., Tenkanen, H. and Toivonen, T. (2015), 'Prospects and challenges for social media data in conservation science', *Environmental Science*, 3, pp. 63

Dunn, P., and Winkler, D. (1999), 'Climate change has affected the breeding date of tree swallows throughout North America', *Proceedings of the Royal Society of London Series Biological Sciences*, 226 (1437), pp. 2487-2490

Dwivedi, Y., *et al*. (2021), 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy', *International Journal of Information Management*, 57, pp. 2-47

Dwyer, R. A., Witte, C., Buss, P., Goosen, W., and Miller, M. (2020), 'Epidemiology of Tuberculosis in Multi-Host Wildlife Systems: Implications for Black (Diceros bicornis) and White (Ceratotherium simum) Rhinoceros', *Frontiers Veterinary Science*, 7, https://doi.org/10.3389/fvets.2020.580476.

Edwards, T., Jones, C., Perkins, S. and Corcoran, P. (2021), 'Passive citizen science: The role of social media in wildlife observations', *PLoS One*, 16(8), e0255416

Edwards, T., Jones, C.B. and Corcoran, P. (2022)' Identifying wildlife observations on twitter', *Ecological Informatics*, *67*, p.101500.

Ehrenfeld, D., (2003), 'Globalisation: effects on biodiversity, environment and society' *Conservation and Society*, pp.99-111.

Eksteen, J. and Bornmann, J (1990), 'Analysis of natural licks at Loskop Dam and Suikerbosrand Nature Reserves', *South African Journal of Wildlife Research*, 20(3), pp. 94-99

Elish, M. and Boyd, D. (2018), 'Situating methods in the magic of Big Data and AI', *Communication monographs*, 85(10), pp. 57-80

Encarnação, J., Baptista, V., Teodósio, M. and Morais, P. (2021), 'Low-cost citizen science effectively monitors the rapid expansion of a marine invasive species', *Frontiers in Environmental Science*, doi: 10.3389/fenvs.2021.725705

EPA (2022), 'Global Greenhouse Gas Emissions Data', *Greenhouse Gas Emissions*, Available at: https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data (Accessed: Nov 2022)

Fabricius, C. (2013), 'The fundamentals of community-based resource management', *Rights Resources and Rural Development: Community-Based Natural Resource Management in Southern Africa.* UK:Routledge, pg. 3-44

Fahrig, L. and Rytwinski, T. (2009), 'Effects of roads on animal abundance: an empirical review and synthesis', *Ecology and Society*, 14

Falzon, G., *et al*. (2019), 'ClassifyMe: a field-scouting software for the identification of wildlife in camera trap images', *Animals*, 10(1), p. 58

Fan, W. and Gordon, M. (2014), 'The power of social media analytics', *Communications of the ACM*, 57 (6), pp. 74-81

Ferreira, S. and Harmse, A. (2014), 'Kruger National Park: tourism development and issues around the management of large numbers of tourists', *Journal of Ecotourism*, DOI: 10.1080/14724049.2014.925907

Field, S., Tyre, A. and Possingham, H. (2005), 'Optimizing allocation of monitoring effort under economic and observational constraints', *Journal of Wildlife Management*, 69(2), pp. 473-482

Flanagin, A. (2017), 'Online social influence and the convergence of mass and interpersonal communication', *Human Communication Research*, 43(4), pp. 450-463

Galaz, V., *et al*. (2009), 'Can web crawlers revolutionise ecological monitoring?', *Frontiers in Ecology and the Environment*, 8, pp. 99-104

Garimella, K., De Francisci Morales, G., Gionis, A. and Mathioudakis, M. (2018), 'Quantifying Controversy on Social Media', *Transactions on Social Computing*, 1(1), pp. 1-2

Garrard, G., Bekessy, S., McCarthy, M. and Wintle, B. (2008), 'When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys', *Austral Ecology*, 33(8), pp. 986-998

Ghaisani, A., Handayani, P. and Munajat, Q. (2017), 'Users' motivations in sharing information on social media', *Procedia Computer Science*, 124, pp. 530-535

Ghermandi, A. and Sinclair, M. (2019), 'Passive crowdsourcing of social media in environmental research: a systematic map', *Global Environment*, 55, pp. 36-47

Gibson, K., et al. (2019), 'Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health', *Scientific Reports*, 9(1), doi: 10.1038/s41598-019-43875-3.

Giovos, I., *et al*. (2016), 'Social media in the service of conservation: a case study of dolphins in the Hellenic seas', *Aquatic Mammals*, 42, pp. 12-19

Goodenough, A. and Hart, A. (2017), *Applied Ecology: Monitoring, managing, and conserving.* Oxford University Press:UK

Goodenough, A., et al. (2017), 'Managing grassland for wildlife: the effects of rotational burning on tick presence and abundance in African savannah habitat', *Wildlife Biology*, 3, doi: https://doi.org/10.2981/wlb.00318

Goodenough, A., Berry, D., Carpenter, W., Dawson, M., Furlong, N., Lamb, R., MacTavish, L., O'Reilly, N., Toms, H., Whitehead, L. and Hart, A. (2023), 'Do you see what I see? Variation in detection, identification and enumeration of mammals during transect surveys', *African Journal of Ecology*, 00, pp. 1-7

Gowdy, J. and Krall, L. (2013), 'The ultrasocial origin of the Anthropocene', *Ecological Economics*, 95, pp. 137-147

Grantham, H., Wilson, K., Moilanen, A., Revelo, T. and Possingham, H. (2009), 'Delaying conservation actions for improved knowledge: how long should we wait?', *Ecology Letters*, 12(4), pp. 293-301

Gruzd, A. and Haythornthwaite, C. (2013), ‘Enabling Community through Social Media’, *Journal of Medical Internet Research*, 15(10), e 248

GSR (2022), *GSR Professional Guidance: Ethical Assurance for Social Behavioural Research in Government*, Crown:UK, Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1000708/2021-GSR_Ethics_Guidance_v3.pdf (Accessed: Jun 2023)

Gray, J. (2018), ‘Challenges of Compassionate Conservation’, *Journal of Applied Animal Welfare Science*, 21, pp.34-42

Greitemeyer, T. (2016), 'Facebook and people's state of self-esteem: The impact of the number of other users' Facebook friends', *Computers in Human Behaviour*, 59, pp. 182-186

Hackel, J. (2001), 'Community Conservation and the Future to Africa's Wildlife', *Conservation Biology*, 13 (4), pp. 726-734

Hák, T., Janoušková, S. and Moldan, B. (2016), 'Sustainable Development Goals: A need for relevant indicators', *Ecological Indicators*, 60, pp. 565-573

Hammer, Ø., Harper, D. and Ryan, P. (2001), 'PAST: Paleontological statistics software package for education and data analysis', *Palaeontologia Electronica*, 4(1), pp. 1-9. Available at: https://www.nhm.uio.no/english/research/resources/past/ (Accessed: Jul 2023)

Hampton, S., *et al*. (2013), ‘Big data and the future of ecology’, *Frontiers in Ecology and the Environment*, 11, pp. 156-16

Hart, A., Carpenter, W., Hlustik-Smith, E., Reed, M. and Goodenough, A. (2018), 'Testing the potential of Twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa', Methods in Ecology and Evolution, 9(11), pp. 2194-2205

Hart, A., Dawson, M., Fourie, R., MacTavish, L. and Goodenough, A. (2022), 'Comparing the effectiveness of camera trapping, driven transects and ad hoc records for surveying nocturnal mammals against a known species assemblage', *Community Ecology*, pp. 1-13

Hausmann, A., Toivonen, T., Slotow, R., Tankenen, H., Moilanen, A., Heikinheimo, V. and Di Minin, E. (2017), ‘Social Media Data Can Be Used to Understand Tourists’ Preferences for Nature-Based Experiences in Protected Areas’, *Conservation Letters*, 11(1), 1-10

Hornsey, M.J. and Fielding, K.S., 2020. Understanding (and reducing) inaction on climate change. *Social Issues and Policy Review*, *14*(1), pp.3-35.

Hofmeester, T., Cromsigt, J., Odden, J., Andrén, H., Kindberg, J. and Linnell, J. (2019), ‘Framing pictures: A conceptual framework to identify and correct for biases in detection probability of camera traps enabling multi-species comparison’, *Ecology and Evolution*, 9(4), pp. 2320-2336

Hsiu-Chia, K. (2018), ‘Social desire or commercial desire? The factors driving social sharing and shopping intentions on social commerce platforms’, *Electronic Commerce Research and Applications*, pp. 1-15

Hu, Y., Manikonda, L. and Kambhampati, S. (2014), 'What we Instagram: A First Analysis of Instagram Photo Content and User Types', *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 8, pp. 595-598

Hudders, L., De Jans, S. and De Veirman, M. (2021), 'The commercialization of social media stars: a literature review and conceptual framework of the strategic use of social media influencers', *International Journal of Advertising*, 40(3), pp. 327-375

Iacona, G., *et al*. (2019), 'Identifying technology solutions to bring conservation into the innovation era', *Frontiers in Ecology and the Environment*, 17(10), pp. 591-598

IPBES (2018): The IPBES regional assessment report on biodiversity and ecosystem services for Africa. Archer, E. Dziba, L., Mulongoy, K. J., Maoela, M. A., and Walters, M. (eds.). Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany. 492 pages. http://doi.org/10.5281/zenodo.3236177

IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Version 1). Zenodo. https://doi.org/10.5281/zenodo.6417333

iSeeMammals (2024), *iSeeMammals*, Available at: https://play.google.com/store/apps/details?id=us.gorges.iseebears&pcampaignid=web_share (Accessed: Mar 2024)

Janecke, B. (2021), 'Mammal Species Richness at Catena and Nearby Waterholes during a Drought, Kruger National Park, South Africa', *Diversity*, 13(8), pp. 387-403

Jarić, I., *et al*. (2021), 'Invasion Culturomics and iEcology', *Conservation Biology*, 35(2), pp. 447-451

Jeffries, M. (2005), 'Extinction', *Biodiversity and Conservation*. 2nd edn. Abingdon, Oxford:Routledge, pg. 12

Jepson, P. and Ladle, R. (2015), 'Nature apps: Waiting for the revolution', *Ambio*, 44, pp. 827-832

Johnston, A., Matechou, E. and Dennis, E. (2022), 'Outstanding challenges and future directions for biodiversity monitoring using citizen science data', *Methods in Ecology and Evolution*, 14, pp. 103-116

Jonason, D., Franzén, M. and Ranius, T. (2014), 'Surveying moths using light traps: effects of weather and time of year', *PloS ONE*, 9(3), pp. 924-953

Kamminga, J., Ayele, E., Meratnia, N. and Havinga, P., 2018. Poaching detection technologies—a survey. *Sensors*, *18*(5), p.1474.

Kasso, M. and Balakrishnan, M. (2013), 'Ex Situ Conservation of Biodiversity with Particular Emphasis to Ethiopia', *International Scholarly Research Notices*, DOI: 10.1155/2013/985037

Kays, R., Crofoot, M., Jetz, W. and Wikelski, M. (2015), 'Terrestrial animal tracking as an eye on life and planet', *Science*, 348 (6420), pp. 247-278

Kearney, M. (2019), 'rtweet: Collecting and analysing Twitter data', *Journal of open source software*, 4(42), pp. 18-29

Kelly, B., Gross, R., Bittinger, K., Sherril-Mix, S., Lewis, J., Collman, R., Bushman, F. and Li, H. (2015), 'Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA', *Bioinformatics*, 31(15), pp/ 2461-2468

Kepe, T. and Hall. R. (2020), 'Land redistribution in South Africa: Towards decolonisation or recolinisation?', *Decolonisation after Democaracy*, Routledge:UK, pg. 128-137

Kindsvater, H., et al. (2018), 'Overcoming the Data Crisis in Biodiversity Conservation', *Trends in Ecology and Evolution,* 9, pp. 676-688

Kitchin, R. (2014), ‘Big Data, new epistemologies and paradigm shifts’, *Big Data Sociology*, pp. 1-12

König, H., et al. (2020), 'Human-wildlife coexistence in a changing world', *Conservation Biology*, 34 (4), pp. 786-794

Kucera, T. and Barret, R. (2011), ‘A History of Camera Trapping’, *Camera Traps in Animal Ecology: Methods and Analyses*. pg.9-26.

Lacea, S. and Ferreira, A. (2014), 'Kruger National Park: Tourism development and issues around the management of large numbers of tourists', *Journal of Ecotourism*, DOI: 10.1080/14724049.2014.925907

Ladle, R., Correia, R., Do, Y., Joo, G., Malhado, A., Proulx, R., Roberge, J. and Jepson, P. (2016), ‘Conservation culturomics’, *Frontiers in Ecology and the Environment*, 14(5), pp. 269-275

Latest Sightings (2023), a, 'Latest Sightings - Connecting People to Wildlife', *Latest Sightings*, Available at: https://latestsightings.com/about-us (Accessed: Aug 2023)

Latest Sightings (2023), b, 'APP', *Latest Sightings*, Available at: https://latestsightings.com/app (Accessed: Aug 2023)

Leadbetter, C. (2008), *We-Think: Mass innovation, not mass production*, London: Profile Books

Leighton, G., Hugo, P., Roulin, A. and Amar, A. (2016), ‘Just Google It: assessing the use of Google Images to describe geographical variation in visible traits of organisms’, *Methods in Ecology and Evolution*, 7(9), pp. 1060-1070

Leung, F., Gu, F. and Palmatier, R. (2022), 'Online influencer marketing', *Journal of the Academy of Marketing Science*, 50, pp. 226-251

Little, S., Harcourt, R. and Clevenger, A. (2001), 'Do wildlife passages act as prey-traps?', *Biological Conservation*, 107(2), pp. 135-145

Liu, L., Mou, C. and Xu, F. (2024), ‘Improved Wildlife Recognition through Fusing Camera Trap Images and Temporal Data’, *Diversity*, 16(3), p. 139

Llausás, A. and Nogué, J. (2012), ‘Indicators of landscape fragmentation: the case for combining ecological indices and the perceptive approach’, *Ecological Indicators*, 15, pp. 85-91

Lorpiete, M., Panzarasa, P., Puliga, M. and Riccaboni, M. (2021), ‘Early warnings of COVID-19 outbreaks across Erope from social medai’, *Scientific Reports, 11(1), p. 2147*

Lowe, J. and Walker, M. (2015), a, ‘Geomorphological Evidence’, Reconstructing Quaternary Environments. 3rd edn. Abingdon, Oxon: Routledge, pg. 19-93

Madon, B., Gimenez, O., McArdle, B., Baker, S. and Garrigue, C. (2011), ‘A new method for estimating animal abundance with two sources of data in capture-recapture studies’, *Methods in Ecology and Evolution*, 2(3), pp. 390-400

Mangachena, J. and Pickering, C. (2021), ‘Why are some animals popular with wildlife tourists: insights from South Africa’, *Journal of Ecotourism*, 22(2), pp. 312-328

Manley, P. *et al*. (2004), ‘Evaluation of a multiple-species approach to monitoring species at the ecoregional scale’, *Ecological Applications*, 14(1), pp. 296-310

Marini, A., Facon, J. and Koerich, A. (2013), 'Bird species classification based on color features', *IEEE International Conference on systems, man, and cybernetics*, pp. 4336-4341

Marini, A., Turatti, A., Britto, A. and Koerich, L. (2015), 'Visual and acoustic identification of bird species', *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2309-2313

Margatyan, L. and Wall-Reinius, S. (2017), 'Commercializing the unpredictable: Perspectives from wildlife watching tourism entrepreneurs in Sweden', *Human Dimensions of Wildlife*, 22(5), pp. 406-421

Marschall, S. (2014), 'Public holidays as lieux de mémoire: nation-buildling and the politics of public memory in South Africa', *Anthropology Southern Africa*, 36 (2), pp. 11-21

McKinley, D., *et al*. (2017), 'Citizen science can improve conservation science, natural resource management, and environmental protection', *Biological Conservation*, 208, pp. 15-28

McShae, W., Velez, J., Shamon, H., Castiblanco-Camacho, P., Tabak, M., Chalmers, C., Fergus, P. and Fieberg, J. (2023), 'An evaluation of platforms for processing camera-trap data using aritificial intelligence', *Methods in Ecology and Evolution*, 14(2), pp. 459-477

Merckx, T. and Slade, E. (2014), 'Macro-moth families differ in their attraction to light: implications for light-trap monitoring programmes', *Insect Conservation and Diversity*, 7(5), pp. 453-461

Mills, D., Fattebert, J., Hunter, L. and Slotow, R. (2019), 'Maximising camera trap data: Using attractants to improve detection of elusive species in multi-species surveys', *PLoS One*, 14(5)

Minin, E., Laitila, J., Montesino-Pouzols, F., Leader-Williams, N., Slotow, R., Goodman, G., Conway, A. and Moilanen, A. (2015), 'Identification of policies for a sustainable legal trade in rhinoceros horn based on population projection and socioeconomic models', *Conservation Biology*, 29(2), pp. 545-555

Montesh, M. (2013), 'Rhino Poaching: A new form of organised crime', *Rhino Resource Centre*, 27, pp. 1-23

Nikodinoska, N., Foxcroft, L., Rouget, M., Paletto, A. and Notaro, S. (2014), 'Tourists' perceptions and willingness to pay for the control of Opuntia stricta invasion in protected areas: A case study of South Africa', Koedoe - *African Protected Area Conservation and Science*, DOI: 10.4102/koedoe.v56i1.1214

Nikolova, A. (2017), 'Integrated approaches for sustainable development goals planning: the case of Goal 6 on water sanitation', *UNESCAP*

Norouzzadeh, M., *et al*. (2018), 'Automatically identifying, counting, and describing wild animals in camera trap images with deep learning', *Proceedings of the National Academy of Sciences*, 115 (25), pp. 5716-5725

O'Connel, A., Nichols, J. and Karanth, K. (2011), 'Science, Conservation, and Camera Traps', *Camera Traps in Animal Ecology: Methods and Analyses*, London:Springer, pg.45-56

O'Conner, J., Smith, M. and James, M. (2017), 'Cameras and settings for aerial surveys in the geosciences: Optimising image data', *Progress in Physical Geography*, 41(3), pp. 325-344

Ozimek, P. and Bierhoff, H-W. (2016), 'Facebook use depending on age: The influence of social comparisons', *Computers in Human Behaviour*, 61, pp. 271-279

Page, M., *et al*. (2020), 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 372, doi: 10.1136/bmj.n71

Panter, C. and Amar, A. (2022), 'Using web-sourced photographs to examine temporal patterns in sex-specific diet of a highly sexually dimorphic raptor', *Royal Society of Open Science*, 9(10)

Parwani, P., et al. (2019), 'Understanding social media: opportunities for cardiovascular medicine', *Journal of the American College of Cardiology*, 73(9), pp. 1089-1093

Pech, D., Condal, A., Bourget, E. and Ardisson, P. (2004), ‘Abundance estimation of rocky shore invertebrates at small spatial scale by high-resolution digital photography and digital image analysis’, *Journal of Experimental Marine Biology and Ecology*, 299(2), pp. 185-199

Pepper, D. (1996), 'Defining Environmentalism', *Modern Environmentalism*, First Edition, London:Routledge, pg. 1-37

Pernat, N., *et al*. (2024), ‘Overcoming biodiversity blindness: Secondary data in primary citizen science observations’, *Ecological Solutions and Evidence*, 5, e12295

Petriello, M.A., Redmore, L., Sène-Harper, A. and Katju, D., 2021. Terms of empowerment: of conservation or communities?. *Oryx*, 55(2), pp.255-261.

Pierson, N. (2021), 'A Multi-Planet Species: Ethical and Environmental Impacts of Privatized Space Travel', *SCIENTIA MORALITAS - International Journal of Multidisciplinary Research*, 8(3), pp. 71-76

Pijanowski, B., *et al*. (2011), ‘Soundscape ecology: the science of sound in the landscape’, *Bioscience*, 61, 203-216

Pittman, M. and Reich, B. (2016), 'Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words', *Computers in Human Behaviour*, 62, pp. 155-167

Pollock, K. and Kendell, W. (1987), 'Visibility bias in aerial surveys: a review of estimation procedures', *The Journal of Wildlife Management*, pp. 502-510

Pollock, K., Marsh, H., Lawler, i. and Alldredge, M. (2006), 'Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs', *The Journal of Wildlife Management*, 70(1), pp. 255-262

Popov, V., Shah, P., Runting, R.K. and Rhodes, J.R., 2022. Managing risk and uncertainty in systematic conservation planning with insufficient information. *Methods in Ecology and Evolution*, *13*(1), pp.230-242.

Powers, J., Carlo, T., Slade, E. and Slik, F. (2021), ‘Biotropica announces a new paper category: Natural History Field Notes’, *bioTROPICA*, 53(2), pp.352-353. doi: 10.1111/btp.12944

Probert, A., Wegmann, D., Volery, L., Adriaens, T., Bakiu, R., Bertolino, S., Essl, F., Gervasinii, E., Groom, Q., Latombe, G. and Marisavljevic, D. (2022), 'Identifying, reducing, and communicating uncertainty in community science: a focus on alien species', *Biological Invasions*, 24(11), pp. 3395-3421

Proulx, R., Massicotte, P. and Pepino, M. (2014), ‘Googling trends in conservation biology’, *Conservation Biology*, 28, pp. 44-51

Reed, M. (2020), ‘Scientific citizens, smartphones and social media-reshaping the socio-spatial networks of participation: Insects, soil and food’, *Moravian Geographical Reports*, 28, pp. 61-67

Ridout, M. and Linkie, M. (2009), ‘Estimating overlap of daily activity patterns for camera trap data’, *Journal of Agricultural, Biological, and Environmental Statistics,* 14(3), pp. 322-327

Roocroft, T. (2012), 'Latest Census & Animal Numbers in the Kruger National Park', Available at: http://www.thekruger.com/stories/censusanimalcount.htm (Accessed: Aug 2023)

Rotman, J. D. et al. (2021) 'Advances in Consumer Research', *Paper #1: Moral Peacocking: Outrage and Identity Signaling on Social Media*, Association for Consumer Research, pp. 786–786

Roulet, J., et al. (2017), ‘Reconsidering the Value of Covert Research: The Role of Ambiguous Consent in Participation Observation’, *Organizational Research Methods*, 20(3), pp. 2-61

Rovero, F., et al. (2008), 'A new species of giant shengi or elephant-shrew (genus Rhynchocyon) highlights the exception biodiversity of the Udzungwa Mountains of Tanzania', *Journal of Zoology*, 274, pp. 126-133

Rowan, A. (2023), 'Rhino Conservation in Africa', *Wellbeing News*, 5(1), pp.2-5

Rowcliffe, M. and Carbone, C. (2008), 'Surveys using camera traps: Are we looking to a brighter future?', *Animal Conservation*, 11, pp. 185-186

Rudel, R., Defries, R., Asner, G. and Laurance, W. (2009), 'Changing drivers of deforestation and new opportunities for conservation', *Conservation Biology*, 23(6), pp. 1396-1405

Safford, H. and Harrison, S. (2001), 'Grazing and substrate interact to affect native vs. exotic diversity in roadside grasslands', *Ecological Applications*, 11(4), pp. 1112-1122

Salehan, M. and Negahban, A. (2013), 'Social networking on smartphones: When mobile phones become addictive', *Computers in Human Behaviour*, 29(6), pp. 2632-2639

SANParks (2007), 'Most Visible Species Checklist', *SANParks Kruger*, Available at: https://www.sanparks.org/parks/kruger/conservation/ff/lists/mammals.pdf (Accessed: Aug 2023)

Santos, J. W., Correia, R.A, Malhado, A.C.M., Campos-Silva, J.V., Teles, D., Jepson, P., and Ladle, R.J., 'Drivers of taxonomic bias in conservation research: A global analysis of terrestrial mammals', *Animal Conservation*, 23 (6) (2020), pp. 679-688

Save the Rhino (2023), 'Poaching Statistics', *Save the Rhino*, Available at: https://www.savetherhino.org/rhino-info/poaching-stats/ (Accessed: Aug 2023)

Sbragaglia, V., Espasandín Soneira, L., Steenbeek, J., Soacha Godoy, K., Piera, J. and Coll, M. (2022), 'The emergence of iEcology and conservation culturomics for the sustainable development of our oceans'.

Scanfeld, D., Scanfeld, V., and Larson, E. (2010), 'Dissemination of health information through social networks: Twitter and antibiotics', *American Journal of Infection Control*, 38(3), pp. 182-188

Schöttker, O. and Wätzold, F. (2022), 'Climate Change and Cost-Effective Governance Mode for Biodiversity Conservation', *Environmental and Resource Economics*, 82, pp. 409-436

Seeley, M. and Asner, G. (2021), 'Imaging spectroscopy for conservation applications', *Remote Sensing*, 13(2), p. 292

Seidlitz, A., Bryant, K., Armstrong, N., Calver, M. and Wayne, A. (2021), 'Signs surveys can be more efficient and cost effective than driven transects and camera trapping: A comparison of detection methods for a small elusive mammal, the numbat (*Mymecobius fasciatus*), *Wildlife Research*, 48, pp. 491-500

Shahbaznezhad, H., Dolan, R. and Rashidirad, M. (2022), 'The Role of Social Media Content Format and Platform in Users' Engagement Behaviour', *Journal of Interactive Marketing*, 52, pp. 47-65

Signorini, A., Maria Segre, A., and Polgreen, P. (2011), 'The use of Twitter to track levels of disease activity and public concern in the U.S. during the Influence A H1N1 pandemic', *PLoS ONE*, 6(5), doi: 10.1371/journal.pone.0019467

Sirami, C., Jacobs, D. and Cumming, G. (2013), 'Artificial wetlands and surrounding habitats provide important foraging habitat for bats in agricultural landscapes in the Western Cape, South Africa', *Biological Conservation*, 164, pp. 30-38

Siyabona Africa (2023), 'Safari Lodges in Kruger National Park', *Siyabona Africa*, Available at: https://www.krugerpark.co.za/Kruger_Park_Private_Concession_Lodging-Travel/Kruger_Park_Private_Concession_Lodging.html (Accessed: Aug 2023)

Sloan, L., Morgan, J., Burnap, P., and Williams, M. (2015), 'Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data', *PloS one*, DOI: 10.1371/journal.pone.0115545

Smith, P., et al. (2014), 'Agriculture, Forestry, and other Land Use (OFOLU)', *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge:Cambridge University Press, pp. 811-922

Smith, D., Smith, Y., Ramesh, T. and Downs, C. (2017), 'Camera-trap data elucidate habitat requirements and conservation threats to an endangered forest specialist, the spotted ground thrush (*Zoothera guttata*)', *Forest Ecology and Management*, 400, pp. 523-530

Soriano-Redondo, A., Bearhop, S., Lock, L., Votier, S. and Hilton, G. (2017) 'Internet-based monitoring of public perception of conservation', *Biological Conservation*, 206, pp. 304-309

Spicker, P. (2011), 'Ethical covert research', *Sociology*, 45, pp. 118-133

Statista (2023), 'Most popular social networks worldwide as of January 2023, ranked by number of monthly active users', *Statista*, Available at: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (Accessed: Aug 2023)

Steffen, W., Grinevald, J., Crutzen, P. and McNeill, J. (2011), 'The Anthropocene: conceptual and historical persectives', *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 369(1938), pp. 842-867

Steger, C., Butt, B. and Hooten, M. (2017), 'Safari Science: assessing the reliability of citizen science data for wildlife surveys', *Journal of Applied Ecology*, 54(6), pp. 2053-2062

Sun, C., Hurst, J. and Fuller, A. (2021), 'Citizen science data collection for integrated wildlife population analyses', *Frontiers in Ecology and Evolution*, 9, pp. 6821-6824

Tafesse, W. and Wood, B. (2021), 'Followers' engagement with Instagram influencers: The role of influencers' content and engagement strategy', *Journal of Retailing and Consumer Services*, 58, pp. 969-989

Tankanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L. and Toivonen, T. (2017), 'Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas', *Scientific Reports*, 7

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H. and Di Minin, E. (2019), 'Social media data for conservation science: A methodological overview', *Biological Conservation*, 233, pp. 298-315

Trolle, M. and Kery, M. (2005), 'Camera-trap study of ocelot and other secretive mammals in the northern Pantanal', *Mammalia*, 69, pp. 409-416

Trolliet, F., Huynen, M., Vermeulen, C., Hambuckers, A. (2014), 'Use of camera traps for wildlife studies. A review.', *Biotechnology, Agronomy, Society and Environment (BASE)*, 18 (3), pp. 446-454

Turner, S. (2013), 'Community-based natural resource management and rural livelihoods', *Rights Resources and Rural Development: Community-Based Natural Resource Management in Southern Africa*, Routledge:UK, pg. 45-66

UNEP (2021), 'Appendices I, II, and III', *Convention on International Trade in Endangered Species of Wild Fauna and Flora*, Available at: https://cites.org/sites/default/files/eng/app/2023/E-Appendices-2023-05-21.pdf (Accessed: Aug 2023)

University of Gloucestershire (2022), *Research Ethics: A Handbook of Principles and Procedures, Gloucestershire*, UK:University of Gloucestershire

UNTWO (2019), 'World Tourism organizations welcomes $1 million donation to grow wildlife tourism for the benefit of people and planet', *World Tourism Organization*, Available at: https://www.unwto.org/world-tourism-organization-welcomes-1-million-donation-grow-wildlife-touris (Accessed: 7/12/2022)

Van der Merwe, P., Saayman, M., Els, J. and Saayman, A. (2017), 'The economic significance of lion breeding operations in the South African Wildlife Industry', *International Journal of Biodiversity and Conservation*, 9(11), pp.314-322

Van Uhm, D.P., 2012. Organised crime in the wildlife trade. *Centre for Information and Research on Organised Crime Newsletter*, *10*(2), pp.2-4.

Vermeulen, C., Lejeuene, P., Lisein, J., Sawadogo, P. and Bouche, P. (2013), 'Unmanned aerial survey of elephants', *PloS ONE*, 8(2), pp. 547-560

Waldron, A., Mooers, A., Miller, D., Nibbelink, M., Redding, D., Kuhn, T., Roberts, J. and Gittleman, J. (2013), 'Targeting global conservation funding to limit immediate biodiversity declines', *Proceedings of the National Academy of Sciences*, 110(29), pp. 12144-12148

Wallach, A., Beckoff, M., Nelson, M. and Ramp, D (2015), 'Promoting predators and compassionate conservation', *Conservation Biology*, 29(5), pp,1481-1484

Wallach, A., Beckoff, M., Batavia, C., Nelson, M. and Ramp, D. (2018), 'Summoning compassion to address the challenges of conservation', *Conservation Biology*, 32(6), pp.1255-1265

Ward, C.D. and Shackleton, C.M. (2016). Natural resource use, incomes, and poverty along the rural–urban continuum of two medium-sized, South African towns. *World Development*, *78*, pp.80-93.

Willemen, L., Cottam, A., Drakou, E. and Burges, N. (2015), 'Using social media to measure the contribution of Red List species to the nature-based tourism potential of African protected areas', *PLoS One*, 10

Willi, M., Pitman, R., Cardoso, A., Locke, C., Swanson, A., Boyer, A., Veldthuis, M. and Fortson, L. (2019), 'Identifying animal species in camera trap images using deep learning and citizen science', *Methods in Ecology and Evolution*, 10(1), pp. 80-91

Wood, S., Guerry, A., Silver, J. and Lacayo, M. (2013), 'Using social media to quantify nature-based tourism and recreation', *Scientific Reports*, 3(1), pp. 2967

Wray, M. (2023), 'An estimated 1600 Lions in Kruger Park', Available at: https://www.krugerpark.co.za/krugerpark-times-3-13-krugers-lions-23272.html (Accessed: Aug 2023)

X (2023), 'Upload Media', *X Developer Platform*, Available at: https://developer.twitter.com/en/docs/twitter-api/v1/media/upload-media/uploading-media/media-best-practices#:~:text=Recommended%20Minimum%20Video%20Bitrate%3A%205%2C000,%2C%201%3A1%20(square) (Accessed: Aug 2023)

Xu, L., Guan, J., Lau, W. and Xiao, Y. (2016), ' An overview of the Pangolin Trade in China', *Traffic Briefing*, pg. 1-10

Yam, J., Bogema, D., Bryant, B., Campbell-Ward, M., Gestier, S., and Jenkins, C. (2018), 'The identification of Theileria bicornis in captive rhinoceros in Australia', *International Journal for Parasitology: Parasites and Wildlife*, 7(1), pp.85-89

Zhuang, H., *et al*. (2022), 'A flagship species-based approach to efficient, cost-effective biodiversity conservation in the Qinling Mountains, China', *Journal of Environmental Management*, doi: 10.1016/j.jenvman.2021.114388

Zijp, M., *et al*. (2017), 'Identification and ranking of environmental threats with ecosystem vulnerability distributions', *Scientific Reports*, 7, p. 9298

Appendix A

Additional Figures

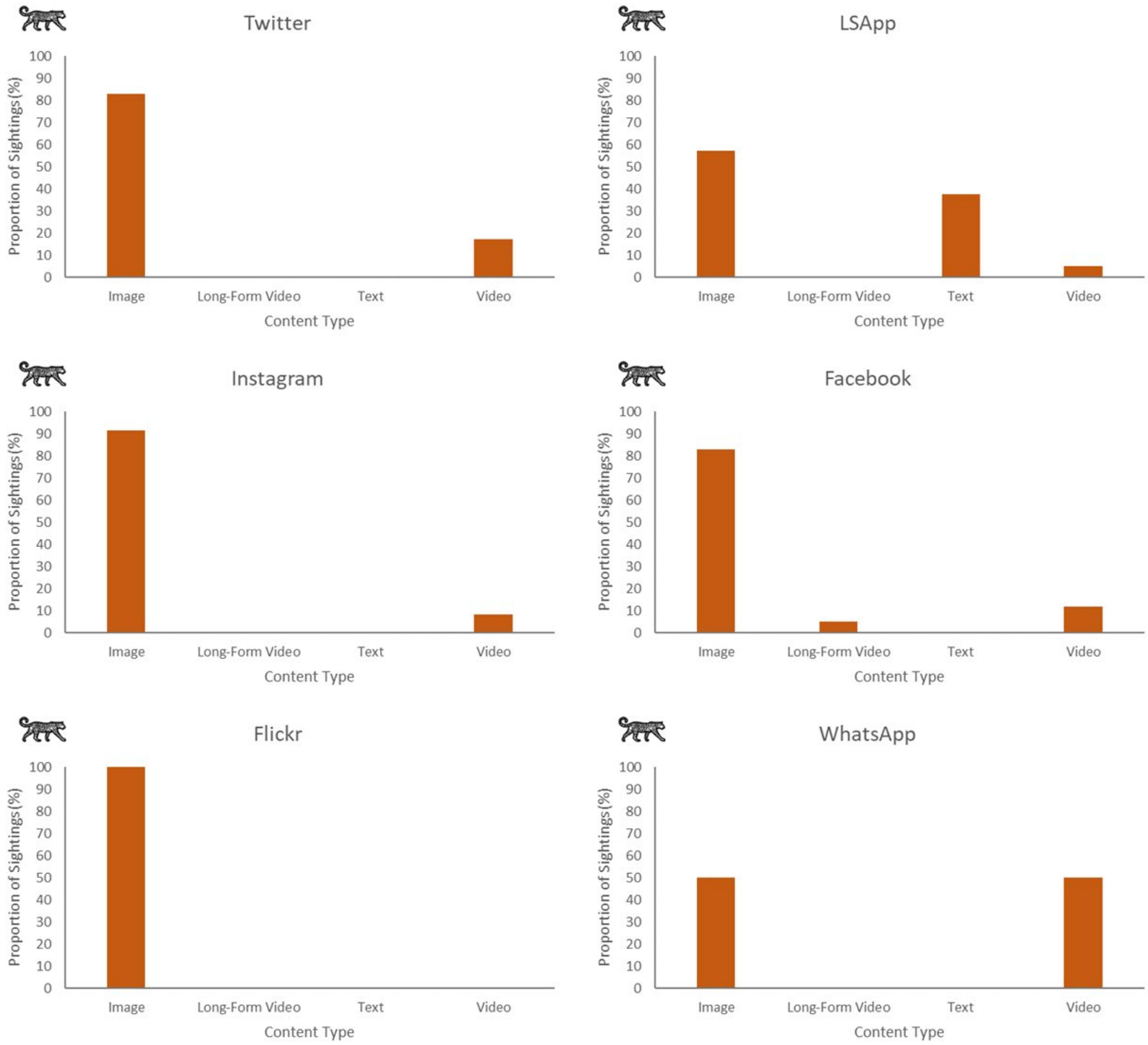

Figure A1: The distribution of leopard sightings by content type across platforms. KNP and PNP pooled. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

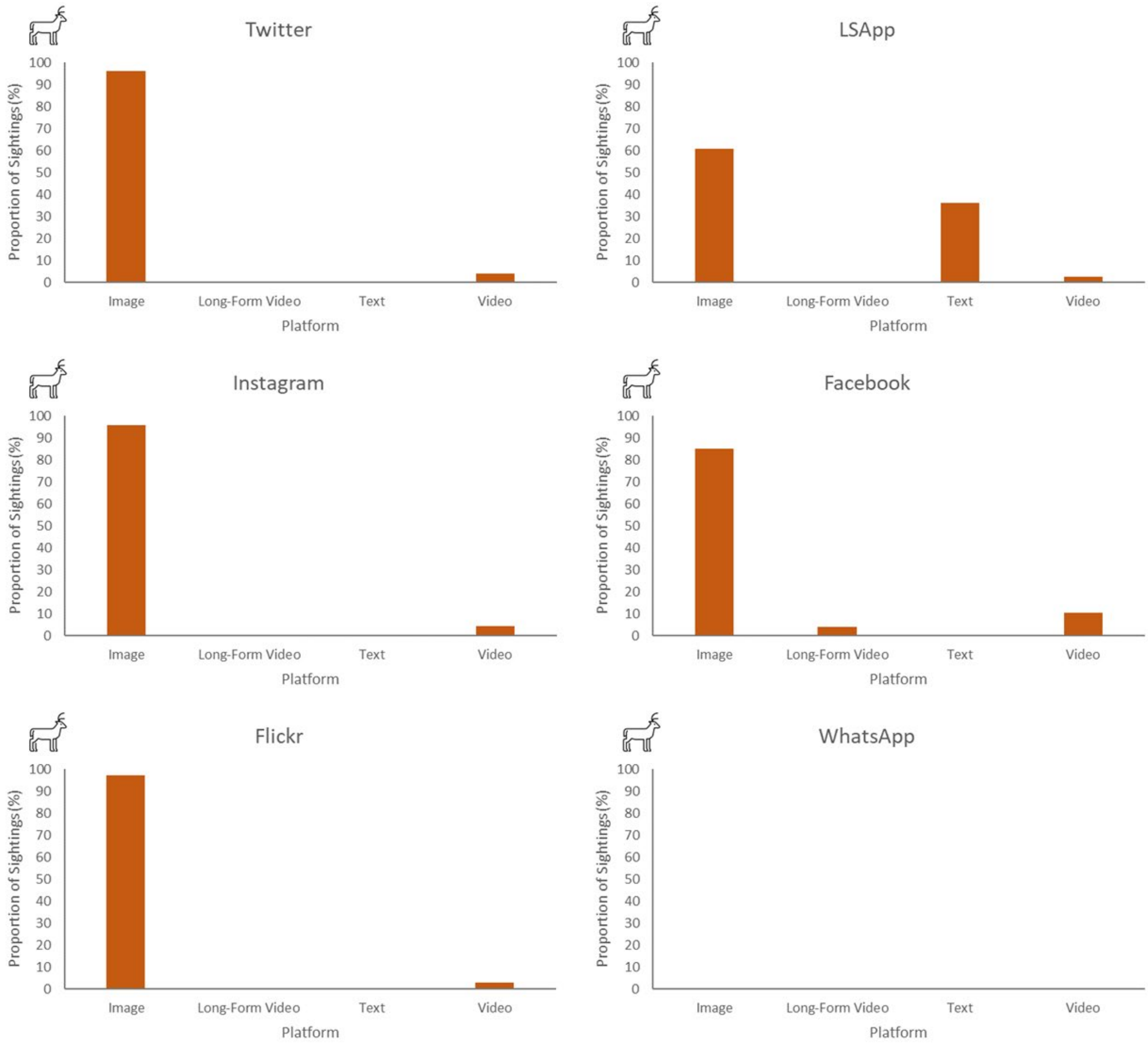

Figure A2: The distribution of antelope sightings by content type across platforms. KNP and PNP pooled. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

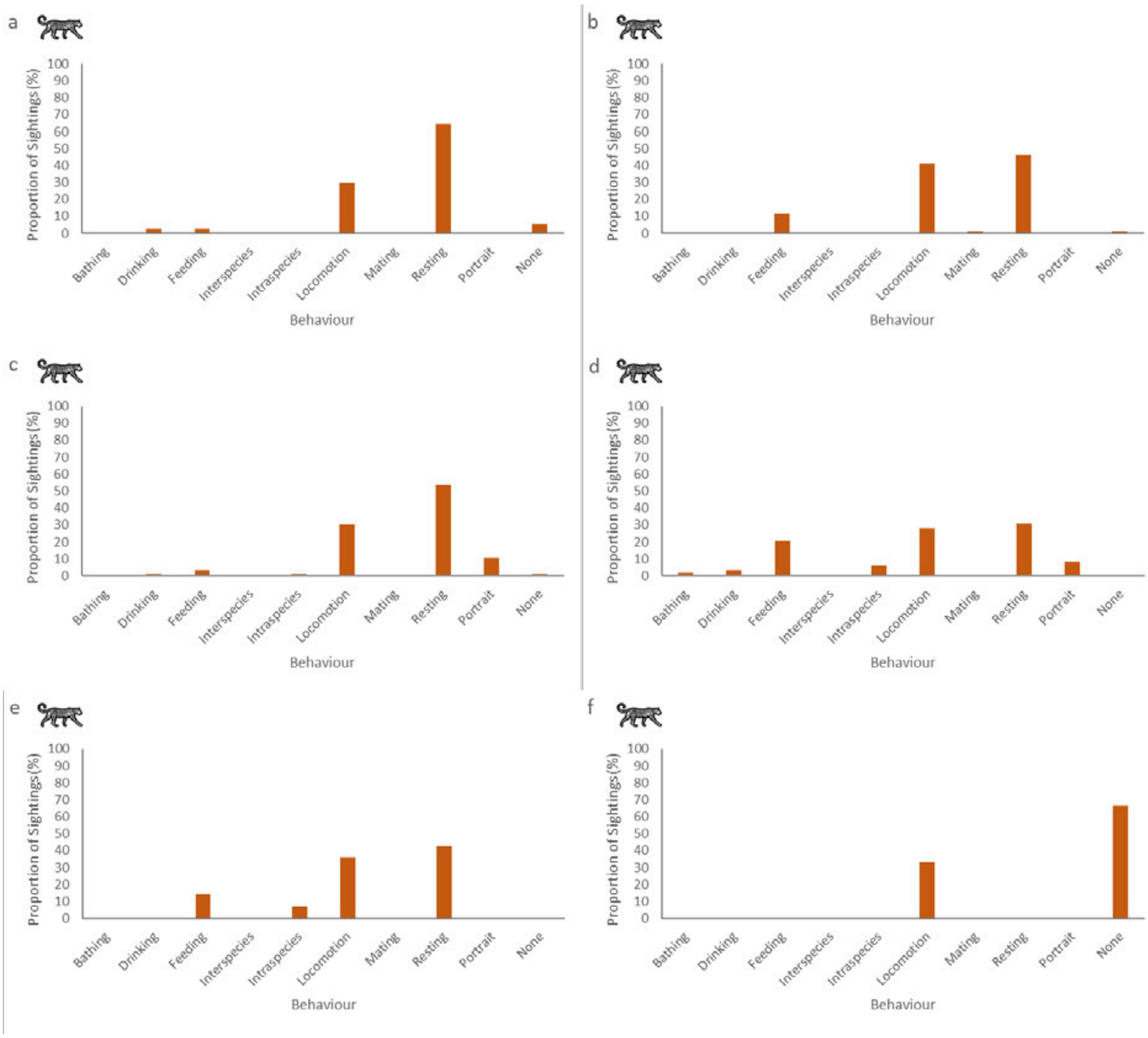

Figure A3: The distribution of leopard sightings on platforms based on behaviour, represented as a proportion of total species sightings on that platform across both parks. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp.

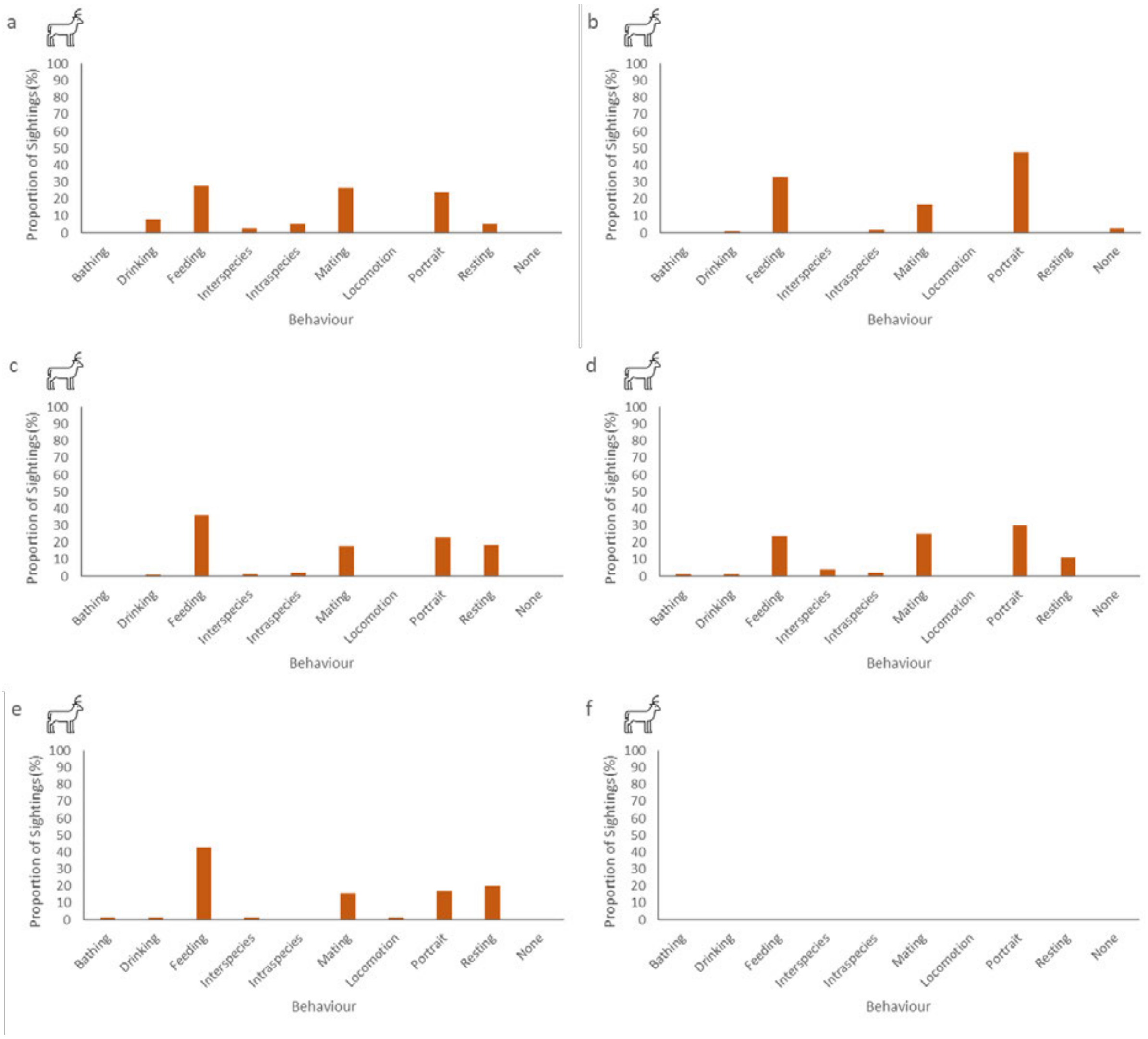

Figure A4: The distribution of antelope sightings on platforms based on behaviour, represented as a proportion of total taxa sightings on that platform across both parks. (a) Twitter, (b) LSApp, (c) Instagram, (d) Facebook, (e) Flickr, (f) WhatsApp

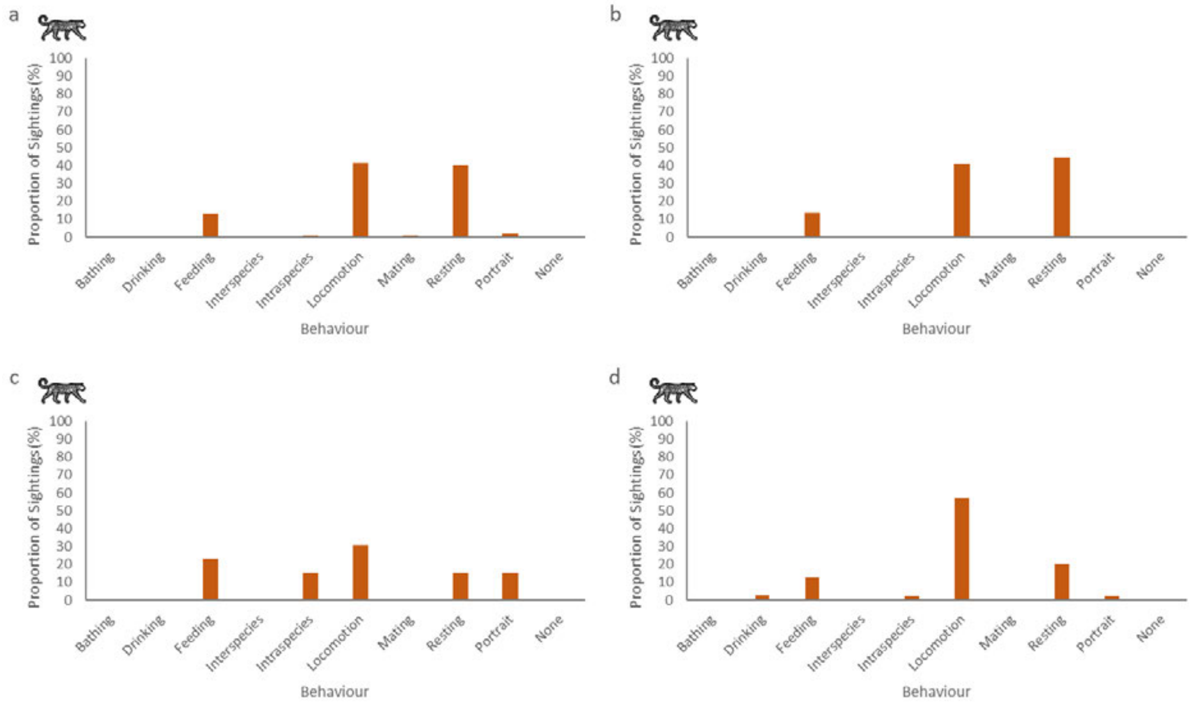

Figure A5: distribution of leopard sightings based on the expression of their behaviour through content type across all platforms and both parks. (a) image, (b) text, (c) long-form video, (d) video.

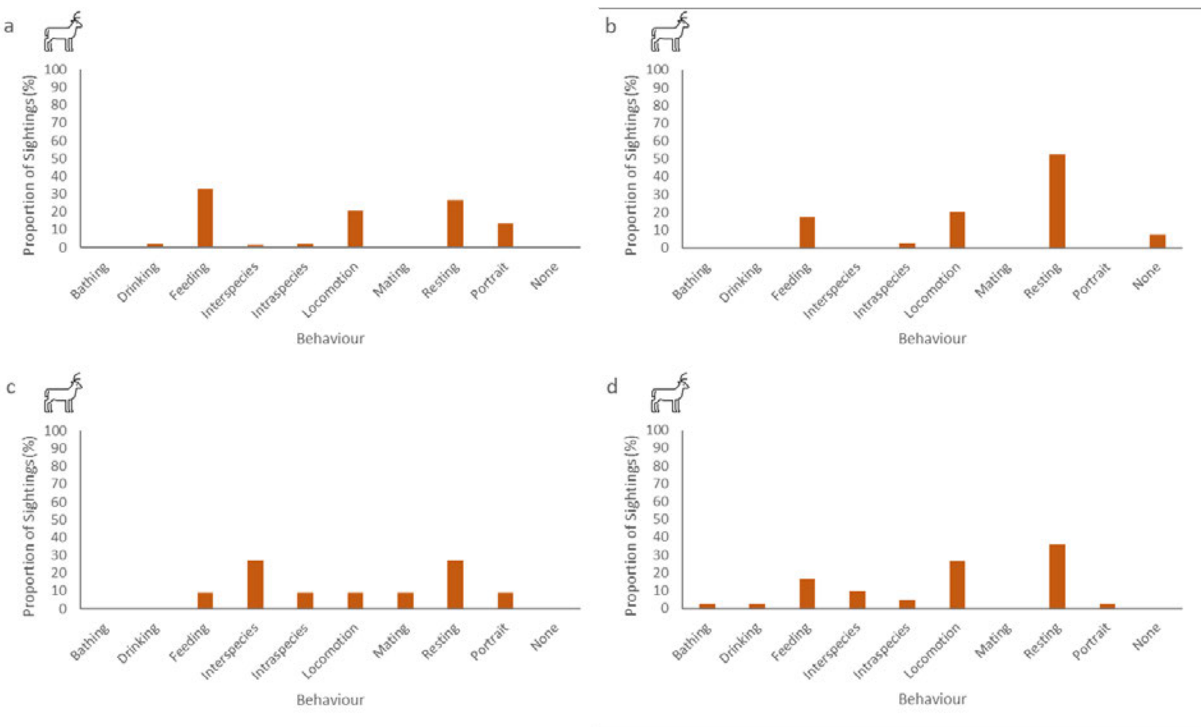

Figure A6: distribution of antelope sightings based on the expression of their behaviour through content type across all platforms and both parks. (a) image, (b) text, (c) long-form video, (d) video.

Appendix B

Alternative Figures

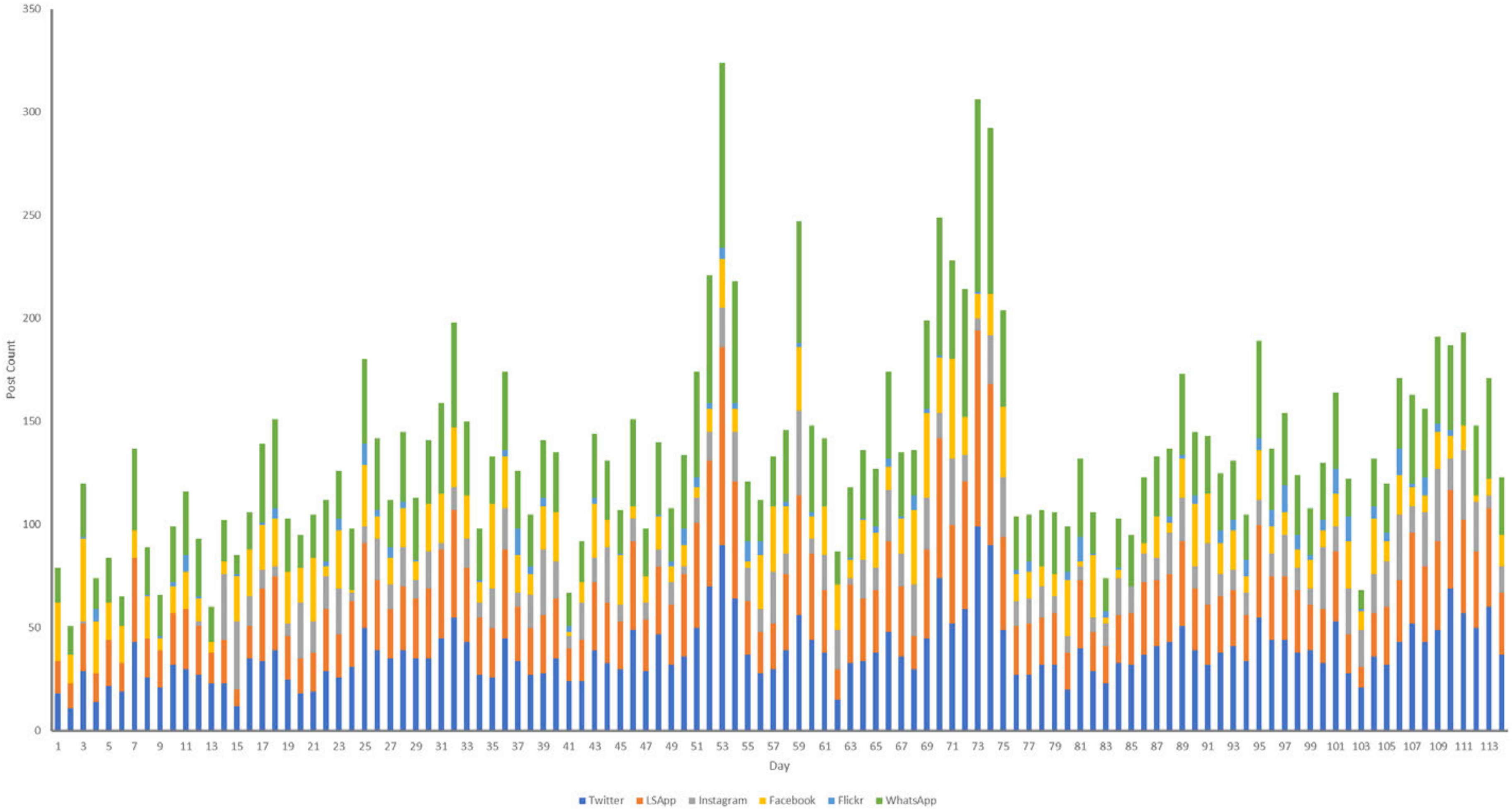

Figure B1: Alternate colouration of figure 4: The number of mammal -sighting-based posts on each platform daily, over the 114-day study period which began on the 16th February 2023 (day 1) and concluded on 9th June 2023 (day 114). Observations were recorded over150 days, but the dataset was reduced to the aforementioned range due to insufficient coverage towards the start and end of this period.

Appendix C

Supporting Documentation

UNIVERSITY OF GLOUCESTERSHIRE

School of Natural, Social & Sport Sciences Research Ethics Panel
Francis Close Hall
Swindon Road
Cheltenham
GL50 4AZ

A The Park
Cheltenham
GL50 2RH
T +44 (0) 3330 14 14 14
W glos.ac.uk

nssethics@glos.ac.uk

Date: 11/08/23

Dear Elliot Tripp

Thank you for your application to the School of Natural, Social & Sport Sciences – School Research Ethics Panel (NSS-SREP).

Following institutional ethical review, I am pleased to confirm ethical clearance.

Please keep a record of this letter as a confirmation of ethical approval for your study (detailed below), reviewed by the School Research Ethics Panel of the School of Natural, Social & Sport Sciences, University of Gloucestershire, on 11th August, 2023

Project Title:	Wildlife Data for Free (WDFF)
Start Date:	01/09/22
Projected Completion Date:	30/10/23
NSS-REP Clearance code:	NSSS.EV1123.01

If you have any questions about ethical clearance, please feel free to contact me. Please use your SREP clearance code in any future correspondence regarding this study.

Kind Regards

[redacted]

IN ANIMO ET VERITATE

Dr Danielle Stephens-Lewis

School Research Ethics Lead

School of Natural, Social & Sport Sciences

Figure C1: Certificate of Ethics Approval for this study, which was initially titled "Wildlife Data for Free", provided by the University of Gloucestershire

WILDLIFE DATA FOR FREE CONVERSATIONS

Hello!
I am Elliot, a Postgraduate by Research student completing my MSc studentship alongside Anne and Adam for my MSc in Environmental Science.

The project that I am completing is called Wildlife Data for Free and aims to (1) determine how posts relating to wildlife vary over different social media platforms and apps; (2) determine what type of content these posts contain; and (3) what potentially useful information such content could harbour in regard to wildlife conservation.

Part of this project includes (hopefully) gaining an understanding of user habits across platforms and inversely, how ecologists and conservationists use social media. I would like to chat with you about these topics, please, via a series of small 20–30-minute conversations.

These conversations will be recorded for future reflection, but all participants will be kept anonymous. They will be guided by prompts which will aim to engage everyone and keep the conversations brief so that you can all enjoy your trips/days. Recordings will be destroyed when the study is concluded in September of 2023.

Participation in these conversations is indicative of consent and understanding.

If you have any concerns or queries about these recordings after the fact, please contact me at

s1804370@glos.ac.uk, or reach out to Prof. Anne Goodenough and/or Prof. Adam Hart either on the trip or upon our return to the UK.

Please print your name below to indicate your understanding of the above information and your willingness to participate in these conversations. Doing so will also aid me in keeping track of the conversations when making notes, dictating recordings, and reviewing the information later on. Again, you will remain **COMPLETELY ANONYMOUS** within the study itself.

Print Name	Signature	Date

Thank you,

Elliot

[redacted]

Figure C2: Focus group consent form to be read and signed by participants. The prelude provided was also read by the researcher at the beginning of each focus group session.