Accuracy, error and bias in species identification

Gail Elizabeth Austen

Durrell Institute of Conservation and Ecology School of
Anthropology and Conservation University of Kent

A dissertation submitted for the degree of
Doctor of Philosophy in Biodiversity Management
2018

Barn owl or apple? ©@teenybiscuit

“The question is not what you look at, but what you see.”

— Henry David Thoreau, *Journal*, 5 August 1851

Acknowledgements

There is no getting away from it, I always have a lot to say, so if you are only looking for your name, you may wish to draw up a chair and pour yourself a drink.

Firstly, enormous thanks go to my ‘super’visory team of Dave Roberts, Markus Bindemann and Richard Griffiths, for believing in me and guiding me through this process. In addition to this role, I’m very grateful for their open door policies, sage advice and good supply of tissues – although not so much Dave, as it was mainly blue roll, which was a bit scratchy. A huge thank you to the University of Kent for the opportunity to get paid to learn under the 50th Anniversary Research Scholarship scheme, and to Richard for introducing me to the concept of PhD while chatting over newts at silly-o’clock on Friday mornings. It is tempting to think of this thesis as the culmination of four years work, but it is the pinnacle of life-long learning. It would be unfair to overlook the contribution made by growing up in a country where compulsory education is often taken for granted, and for those unsung heroes whose dedication to their students really does change lives. A few names spring to mind - Mrs Riddick, Mr Southwick, Mr Gray, Mr Holness, and Mr Marsh, whom I remember most fondly thanks to his kind manner and generous nature.

None of this research would have been possible without the study participants and my gratitude goes to all those that gave up their time to take part in this research. Thanks also go to the Bumblebee Conservation Trust (BBCT) and British Ecological Society (BES) citizen science network for circulating and promoting surveys. I am also grateful to the Amphibian and Reptile Groups of the UK (ARG UK) for allowing me to recruit participants at the Herp Workers’ Meeting. Special thanks are due to those who were generous with their time, support, and advice over the course of this project. This includes Nikki Gammans (BBCT), Stuart Roberts from the Bees Wasps and Ants Recording Society (BWARS), and Paul Williams from the Natural History Museum (NHM), all of whom kindly shared knowledge of their study species. Appreciation also goes to DICE staff members Zoë Davies and Nicola Kerry-Yoxall, who have helped in various ways, but mainly for persuading me that a PhD is achievable. I also wish to acknowledge the importance of peer support. The Room 172 Crew have seen me through some tough times by a variety of means: ridiculous cat videos from Tristan Pett, hilarious capoeira demonstrations from Rob Ward, lots of chats and punches from my Krav buddy Amy Hinsley (no. 3), lattes (but never enough time) with Hazel Jackson, and fighting Janna Steadman for baby cuddles, courtesy of Jinks Robinson (the babies, not the cuddles), have made it all bearable. Oh, and copious amounts of tea.

Where do I start with family? A massive thanks to Mum & Dad for more than I could ever write here, but particularly for showing me that there is a world out there and encouraging me to explore it. Thanks to Steve for being a hard taskmaster at the beginning of this journey (my OU degree). I'm sure that he would have been very proud of me. Thanks to Wendy for her love and support, and to Alex & Maddie, who inspire me to try and make the world a better place, and to lead by example. Also to Si – although a latecomer to this party, his presence is most welcome.

Ultimately, I am thankful for being born into a world of opportunity and wonder. The freedom to ask "How?" and "Why?" without retribution, is a privilege for which I am eternally grateful.

Declaration of Originality

Gail Austen wrote all the chapters within this thesis, with each chapter benefitting from editorial suggestions by PhD supervisors David Roberts, Markus Bindemann, and Richard Griffiths.

Chapter 2: Gail Austen, David Roberts, and Markus Bindemann conceived the idea. Gail Austen conducted the literature review, designed and delivered the experiment, analysed the data, and wrote the manuscript, with input from Markus Bindemann and David Roberts. Richard Griffiths provided advice on analyses.

Chapter 3: Gail Austen, David Roberts, and Markus Bindemann conceived the idea and designed the experiment. Gail Austen conducted the literature review, delivered the experiment, analysed the data, and wrote the manuscript with input and guidance from Markus Bindemann and David Roberts.

Chapter 4: Gail Austen conceived the idea following a presentation at the Herpetofauna Workers' Meeting in 2014. The presenter spoke about a party surveying for newts all stepped over what they believed to be a smooth newt. However, the presenter noted that it was an alpine newt, and on closer inspection everyone agreed on his identification. Gail Austen conducted the literature review, designed and delivered the experiment, analysed the data, and wrote the manuscript with input from David Roberts, Markus Bindemann and Richard Griffiths.

Chapter 5: Gail Austen conceived the idea, conducted the literature review, designed and delivered the experiment, and wrote the manuscript with guidance from Markus

Bindemann and David Roberts. Professor Martin Ridout (School of Mathematics, University of Kent) advised on statistical analysis, and provided the bespoke R code to calculate measure of effectiveness (ME) values, and to create the fluctuation diagram.

Abstract

Reliable, robust data is fundamental to effective decision-making. Species observations are used as evidence in a range of areas that work towards conserving biodiversity. Decisions made on these data are only well informed if the species have been accurately identified. Moreover, the *mis*identification of species can have widespread socio-economic impacts. Despite these important applications of species data, the possibility of accuracy, error, and bias in species identification remains largely unexplored. Both volunteers and professionals conduct species identification, and in its simplest form, this process is a judgement made by reference to identification aids, or from prior knowledge. This thesis aims to fill an essential knowledge gap by investigating accuracy in species identification between individuals, across levels of expertise, and the levels of agreement between individuals with similar experience. Applying methods from forensic face recognition research, individuals with varying levels of expertise, and interest in biodiversity, participated in a series of simple image-based tasks. These tasks involved online, pairwise matching tasks under optimised conditions, and sorting tasks with images downloaded from Internet sources. This study shows that decisions on species identification are highly variable between individuals, and high levels of accuracy are achievable by experts *and* non-experts. Moreover, experience is no guarantee of accuracy, and inter-specific disparity does not always exceed intra-specific variation. There is a need for a simple, principled method for assessing identification accuracy, which can be performed by experts and non-experts alike. This method also needs to be sensitive enough to capture individual differences. Improvements in technology have led to an increase in data being collected from previously inaccessible areas, and citizen science has widened participation. However, as data collection adapts to incorporate changes in

how species observations are collected and by whom, methods for assessing and evaluating the reliability of those data must evolve.

Table of Contents

Figures

Tables

The Gruffalo © Julia Donaldson and Axel Scheffler 1999

"But who is this creature with terrible claws
And terrible teeth in his terrible jaws?
He has knobbly knees and turned-out toes
And a poisonous wart at the end of his nose.
His eyes are orange, his tongue is black;
He has purple prickles all over his back."

(Donaldson 1999)

Chapter 1.

1 Introduction

Anyone familiar with the story will recognise ‘this creature’ as the eponymous ‘Gruffalo’ (Donaldson 1999). This story of a mouse portraying a dangerous, mythical creature to survive the perils of the forest demonstrates the power of description. When it subsequently transpires that this creature is in fact real, the mouse continues to use the creature’s presence to its advantage. This description may belong in a storybook or a child’s imagination, but it is worth noting that there is a scientific description for the ‘Loch Ness Monster’ (*Nessiteras rhombopteryx).* Based on a grainy image of a supposed flipper, this was an attempt to protect the species in case it did exist, as protection under UK legislation is not granted without a valid Latin name (Scott & Rines 1975). Paradoxically, another creature that lacks a specimen on which the species was scientifically described and named (known as a holotype) is the modern human (*Homo sapiens*) (Spamer 1999). This process of describing and naming specimens, known as taxonomy, assigns a unique name to a taxon, and species identification is the application of this unique identifier to an unknown organism. Species identification is performed on a daily basis, for example in species counts, monitoring disease outbreaks, and the movement of wildlife across borders. While accurate species identification is fundamental to providing reliable data, avoiding *mis*identification is equally important, as it can have negative socio-economic impacts. Despite these implications, little is known about how well humans perform in this seemingly simple task.

The diversity of life on Earth is well studied, yet the number of species in existence remains unknown. During 250 years of modern taxonomic classification, over 1.2 million species descriptions have been catalogued (Mora et al. 2011), and this process of assigning a unique name to a species means that a universal identifier exists for every *known* species. However, the true number of species is difficult to determine. Estimates include an approximation of 8.7 million species, with 86% of terrestrial species and 91% of marine species yet to be described (Mora et al. 2011), and the inclusion of microbial diversity puts the figure over 1 trillion (10^{12}) (Locey & Lennon 2016). Although taxonomists routinely define and name species, the identification of species is not limited to this specialist community. Species identification is a daily occurrence in a range of professional and non-professional roles, yet the patterns and processes associated with accurate species identification remain largely unexplored. This thesis uses established methodologies from forensic face recognition research to investigate accuracy in species identification, and to determine if there are any associated patterns of error or bias.

1.1 What is species identification and why is it important?

Species identification, as well as species *mis*identification, can have far-reaching consequences. The ability to *recognise* a species, even if we cannot name it, may be sufficient in certain circumstances, for example knowing which wild plants are safe to eat. Some people associate the ability to identify and name an organism with small groups of professionals or enthusiasts, such as taxonomists, keen gardeners or amateur naturalists. However, accurate species identification is crucial in many domains, and misidentification can impact society in a number of ways.

1.1.1 Human health and wellbeing

In terms of human health, species misidentification can affect individuals on different scales, from food consumption to those reliant on ecosystems. When a toxic food source is mistaken for an edible one, such as errors made when foraging (Public Health England 2014, Diaz 2016), it not only causes suffering to the consumer, but also places an unnecessary burden on health services. Accurate species identification by others can also affect humans directly and indirectly. Statutory bodies are often tasked with identifying and monitoring potential dangers to public health, and this information relies on accurate species identification. For example, identifying species of mosquitoes and their distributions is crucial when attempting to prevent and manage malaria outbreaks (Hardy & Barrington 2017). In the UK, *Culex modestus* - found on marshland in Kent - is monitored at all life stages, as it is a potential vector for the West Nile virus (Vaux & Gibson 2015). On a larger scale, humans are wholly reliant upon ecosystem services. The 1982 United Nations (UN) World Charter for Nature and the 1992 Earth Summit in Rio de Janeiro recognised the importance of species and habitats to human wellbeing (Farhan Ferrari, de Jong & Belohrad 2015). Such ecosystems are often monitored and considered 'healthy' by the presence or absence of certain species (Díaz et al. 2006). For example, dinoflagellates and diatoms act as primary producers, but can also create harmful algal blooms (Culverhouse et al. 2003; Hinder et al. 2012). One crucial ecosystem service that is of conservation concern is pollination. With over 90% of flowering plants in terrestrial ecosystems being pollinated by animals (Memmott, Waser & Price 2004), accurate species identification is crucial to understanding the complex interactions between plants and pollinators. Accurate species identification is also important to food security, for example, the assessment and management of fish stocks (Beerkircher, Arocha & Barse 2009; Tillett et al. 2012), and the monitoring of fungal pathogens that cause crops to fail (Crous et al. 2016). In addition to health and food security, accurate species identification can also affect local and national economies (Nellemann et al. 2016).

1.1.2 Economic impacts

Food provision, fish stocks, forestry and agricultural activities are of great economic importance worldwide. For example, the global economic value of pollination was estimated to be €153 billion in 2005 (Gallai et al. 2009). Furthermore, accurate assessments of pollinator populations are required to evaluate how any declines in pollinators could affect agricultural markets, especially for crops that can only be pollinated by single or a few species (Gallai et al. 2009). In addition, the international trade in wildlife has positive and negative impacts on economies across the world. Illegal wildlife trade has negative impacts on livelihoods and economies (Nellemann et al. 2016) and is difficult to value. However, it is thought to be more lucrative than the illegal trade in small arms, human organs, and diamonds (Hinsley, King & Sinovas 2016). Likewise, the *legal* wildlife trade is also lucrative, and can provide employment, support local and national economies, and can be vital to some rural communities (Oldfield 2003). To ensure sustainability of the plants and animals being traded, these markets are regulated, and both legal and illegal trade are monitored and regulated on a global scale. For example, the Convention on International Trade in Endangered Species (CITES) is concerned with over 35,000 species, as listed in the CITES appendices (CITES 2017). To ascertain which specimen is being traded, irrelevant of whether it is a legal or illegal sale, accurate species identification is required. This process involves enforcement officials checking paperwork, or seizing suspected illegal items, on which they need to make a judgment. This could entail identifying a specimen, or a part of one, which they may not have previously encountered. Furthermore, with the increase in illegal trade over the Internet (Harrison, Roberts & Hernandez-Castro 2016), enforcement officials need to identify species from images rather than specimens, and increasingly seeking expert opinion on these images.

Accurate species identification can also have a significant impact on the building and development sector. In 2016, the UK construction industry was valued at nearly £93.5 billion (Office for National Statistics 2017), and many of these infrastructure and housing projects require Environmental Impact Assessments (EIAs) as part of the planning process. These assessments determine the potential impact of a development on the environment, and are referred to by decision makers and the public (Briggs & Hudson 2013). Initial surveys collect data on the presence of certain species and habitats, which in turn determines if further surveys are required (Treweek 2009). The findings from these surveys lead to decisions on whether a project is to be halted, if mitigation measures are to be implemented, or if compensation is to be paid (Briggs & Hudson 2013). While costs are incurred in surveying, irrelevant of the outcomes, accurate species identification is paramount to ensuring that subsequent business decisions are well informed. Halting a development can be costly to businesses and the local economy, plus the cost of survey and mitigation measures can be difficult to budget for, as the expense can vary within and between species. In England alone, the annual cost of mitigating against the impact of development on great crested newts (*Triturus cristatus*) is estimated to be between £20 million and £43 million (Lewis, Griffiths & Wilkinson

2017). However, the failure to monitor species and implement relevant mitigation can also incur financial penalties. When surveying for great crested newts, the animals can only be handled under specific licences granted by the relevant UK authority, as they are a European Protected Species (EPS). Whether observers are aware of this restriction or not, many people choose to photograph rather than handle specimens, as they wish to minimise any unnecessary stress or damage to the organism or its habitat. However, observations made *without* handling a specimen are not always as informative. When a specimen is handled, the observer has the opportunity to view and touch a range of features that may not otherwise be visible to them. As a result, many novice observers seek help with identification from those with experience or expertise, especially for species of conservation concern.

1.1.3 Monitoring and conserving biodiversity

The conservation of biodiversity poses an immense challenge. It requires the use of existing data, the collection of new data, and the monitoring of wildlife on a vast scale (Humphries et al. 1995; Powney & Isaac 2015). Moreover, when monitoring, managing, and sustaining biodiversity, accurate species identification is paramount to make these data informative and reliable (Farnsworth et al. 2013). Species observations provide evidence of species distributions (Evans et al. 2016), intra- and inter-specific interactions, and how species adapt to changing environments (Pimm et al. 2015). Human activities are responsible for unprecedented rates of extinction (Pimm et al. 1995), with overexploitation and agriculture being the biggest drivers of species loss, followed by urban development, invasion, and disease (Maxwell et al. 2016). Investigating these patterns and processes draws on a variety of information, such museum collections (e.g. specimens, label data, illustrations and field notes)

(Hill et al 2012), and the modelling of observation data to predict species distributions over space and time (Hernandez et al. 2006). The outcomes of these studies are reliant on the accuracy of the input data, yet few studies discuss possible error rates, or the identification skills of those collecting the data.

Monitoring takes place for a number of reasons, such as general curiosity, legislative requirements, or to answer specific questions (Lindenmayer & Likens 2010). Emerging technologies have enabled this collection of data to occur in vast quantities, and in some cases beyond traditional human observation (Pimm et al. 2015). An international shift towards evidence-based policy has led to an increase in demand for data to inform decisions relating to the conservation of biodiversity (Baillie, Collen & Amin 2008; Sutherland & Burgman 2015; Turnhout, Lawrence & Turnhout 2016). Such data are used in studies of community ecology (Sih & Christensen 2001), extinction risk (Purvis et al. 2000; Roberts, Elphick & Reed 2010; Boakes, Rout & Collen 2015), biosecurity (Crous et al. 2016), emerging diseases (Mysterud et al. 2016), invasive species (Roy & Brown 2015), monitoring wildlife trade (Phelps & Webb 2015), and predicting impacts from changes in the environment (Jones 2011). Conversely, species for which there are little data remain vulnerable as gaps in information create uncertainty. For example, monitoring population trends (Baillie, Collen & Amin 2008), setting conservation priorities (Rodrigues et al. 2006) and identifying drivers of extinction (Bland & Collen 2016) are problematic without reliable baseline information on the species involved (Boakes et al. 2015). For data to be beneficial to any decision-making process, it must be reliable, robust and accurate. However, observing, identifying and counting organisms can be problematic (Elphick 2008), and species misidentification can have serious negative impacts. Examples of this include the accidental culling of the takahē *Porphyrio hochstetteri*, categorised as endangered by the IUCN Red List (BirdLife International 2016) which was mistaken for the destructive pūkeko *Porphyrio porphyrio melanotus* (Hunt 2015), and public resources being wasted on creating a management plan for the extinct ivory-billed woodpecker (*Campephilus principalis*) following a mistaken identification of this bird (Solow et al. 2012).

As well as the socio-economic impacts outlined here, invasive species can also threaten biodiversity through mechanisms such as predation, competition, and hybridization (Latombe et al. 2016). Managing biological invasions requires accurate information on the distribution of native species, as well as non-native species, in order to monitor the impact (Latombe et al. 2016). However, detecting novel species can be problematic. For example, observer expectations can influence visual perception (Kassin et al. 2013), and a correct identification may be discounted if the specimen has been observed outside of its expected range. There are numerous examples of escapees surviving in the wild, which can often be known locally but not recorded nationally. For example, wallabies (*Macropus* spp.) are occasionally recorded in Kent (see https://youtu.be/vz47ZFCat1g), but the UK’s National Biodiversity Network Atlas (www.nbnatlas.org) has no records for the county. Unusual observations can be referred to specialist individuals or organisations, as either specimens or images. However, people are often encouraged to send images in the first instance, as this may be more practical and reduces risks associated with biosecurity (Government of Western Australia, 2017; Harlequin Ladybird Survey, 2017). If the image proves to be of interest, a specimen or further observations may be sought, but it is still the initial identification of the image that influences this decision.

Species can also be identified through molecular techniques. While these methods are widely used, they are still linked to, and used in conjunction with, morphological descriptions (Renner et al. 2016). For example, DNA barcoding matches a sequence against a reference library of sequences from known species. However, these sequences are only as accurate as the identification of the specimen from which the material was extracted. Furthermore, without a voucher specimen or image, the original identification cannot be visually verified. Genetic methods of identification are informative, especially for species that are very similar in morphology (Williams et al. 2012). For example, the '*lucorum*' complex of bumblebees (*Bombus*) can only be distinguished from their DNA (Scriven et al. 2015). However, molecular techniques are complementary to, rather than a substitute for, morphology (Will & Rubinoff 2004).

1.2 When is a species a species?

Darwin noted that "*No one definition has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species*" (Darwin 1859, pg. 101). Despite species being the basic unit for the study of the natural world (Mallet 2001; Agapow et al. 2004), there is no agreement on how to define what species are, or how we should identify them (Hey 2001; Hey 2006; De Queiroz 2007). Strict rules of nomenclature help stabilise the naming of species (Pyle & Michel 2008), but identifying and naming a species is quite different to defining what a species is (Kunz 2012). John Ray is credited with coining the term 'species' in 1686, taken from the Latin for ‘kind’ (Wilkins 2011). However, there is a perennial lack of agreement over what constitutes a species (Mayr 1942; Mayden 2002; Hey 2001; De Queiroz 2007). Winston (1999) refers to a glut of literature on the ‘species problem’ and comments that it has been “discussed, argued over, and symposiumed to death”. Up to 27 different species concepts have been suggested (Wilkins 2011), difficulties surrounding the species problem is a conflation of the function of species concepts with their application. (Mayden 1997). The process of evolution results in speciation and adaptation, and many, but not all, species concepts relate to how lineages evolve (De Queiroz 2007). In reality, inter-species variation is vague and can overlap. Nevertheless, species delimitation is essential for organisms to be named and studied (Vane-Wright 2000), and continues to be debated. There has been a call for a unified species concept, as contemporary concepts share the common element of separately evolving metapopulation lineages (De Queiroz 2007). However, there is also an argument that biodiversity is a phenotypic concept, and the role of a species is more important than its lineage (Freudenstein et al. 2016).

Species arise by a variety of mechanisms, and different species concepts relate to the different traits of populations (Agapow et al. 2004). Mayr (1942) discusses the dynamic nature of species and that species concepts need to adapt to emerging theories around genetics and evolution. Mayr’s biological species concept, whereby species are defined by their ability (or potential ability) to interbreed, is popular, but does not incorporate organisms that reproduce asexually (Queiroz 2005) or that are extinct. Table 1.1 is taken from Queiroz (2005), summarising the major species concepts in use, and the properties that define them as an alternative species concept.

Advances in genetics have meant that taxonomic designations based solely on morphology need re-examining, and at scales that differ according to the type of organisms being studied (Hendry et al. 2000). The failure to define a single, usable concept is a philosophical debate that is likely to continue (Kunz 2012). However, continuity can be found in the stability of recognising and naming species, and a shared understanding of what that species name means. The publication of a new or reclassified species does not mean that the new name is universally adopted, but apart from a few exceptions (e.g. Hoser 2015), names generally become accepted and synonymy allows any change of species name to be tracked.

Table 1.1. Additional properties frequently considered necessary for separately evolving metapopulation lineages to be deemed a species (From Queiroz (2005), pg. 6603).

Property	**Species concept and/or definition**
Potential interbreeding	Biological species concept definition
Shared specific mate recognition or fertilization	Recognition species concept
Same niche or adaptive zone	Ecological species concept
Monophyly (as inferred from apomorphy or exclusive system coalescence of gene trees)	Monophyly version of the phylogenetic species concept, genealogical species concept
Form a phenetic cluster (quantitative difference)	Phenetic species concept
Form a diagnosable group (fixed qualitative difference)	Diagnosable version of the phylogenetic species concept, some interpretations of the evolutionary species concept
Form a genotypic cluster	Genotypic cluster species definition

The human tendency to generate order and understanding has led to the creation of discrete taxonomic units along what is essentially a continuum of heritable characters (Heywood 1998; Hey 2001). This classification is critical for discussing and studying biological systems, and differences between certain taxa are evident, (e.g. insects and mammals, whereas intra- and inter-species disparities are not always obvious).

Species variation is heterogeneous (Gaston 2000). When a species is described, a ‘type’ specimen is defined for that name, which acts a ‘template’ for that species (Krell 2004). However, evolutionary processes give rise to intra-specific variation, inter-specific similarity, and some species having very different forms depending on their life stage. The practice of species identification involves ‘telling things together’ as well as ‘telling things apart’, but heterogeneity within and between taxa is not standard. Some species undergo metamorphosis, and individuals can differ so much at different life stages that they are virtually unrecognisable as the same organism (Fig 1.1). These sudden and conspicuous changes in development can be found in a range of different taxon groups, for example fish, amphibians, insects, jellyfish and plants. For some species, identification can be made at the egg stage (e.g. birds – Class: Aves), at the larval stage (e.g. the caterpillars of butterflies and moths – Order: Lepidoptera), or using a shed exoskeleton (exuviae) (e.g. dragonflies – Order: Odonata, and ladybirds – Family: Coccinellidae). However, this determination is not possible at all non-adult stages. Some larval stages are not fully visible, such as lacewings (Family: Chrysopidae), which cover themselves with sand grains. Accurate identification may also rely on sexual characteristics that have yet to develop, as seen with newts (Family: Salamandridae). Phenotypic plasticity can also be temporary or reversible in response to changes in the environment, such as morphological changes observed in tadpoles when encountering predators and competitors (Relyea 2004).

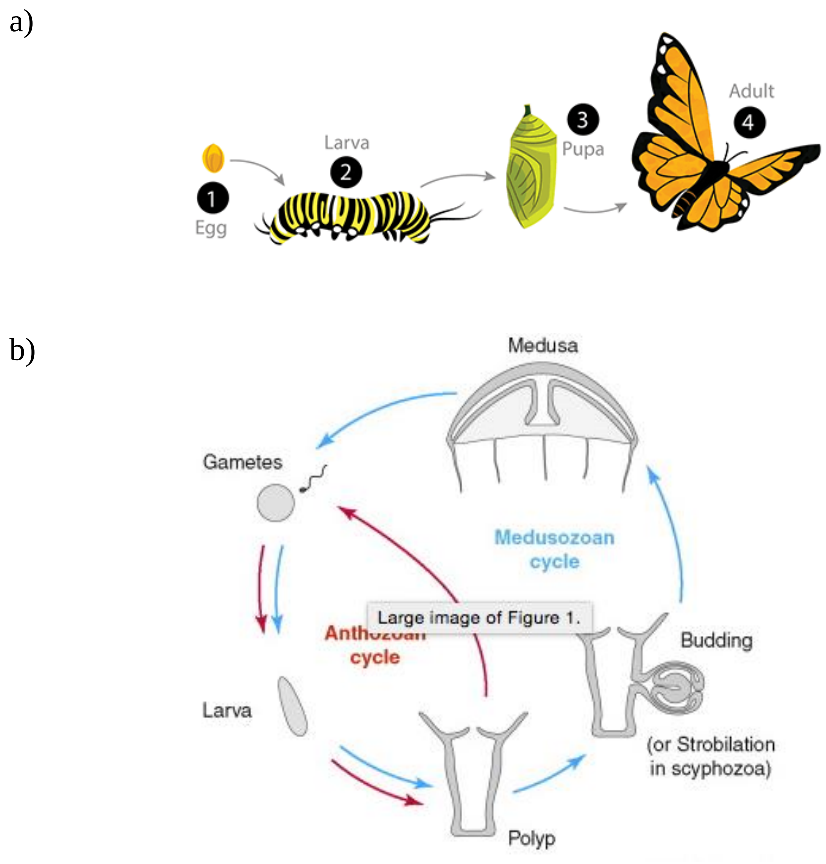


Figure 1.1. Examples of metamorphosis in a) butterflies (Order: Lepidoptera) (askabiologist.asu.edu 2016) and b) jellyfish (Phylum: Cnidaria) (Forêt et al. 2010).

Organisms can also vary within a population, for example polymorphism where individuals differ in appearance despite being the same species. Some individuals have colour variation, such as all white (leucistic), all black (melanistic), or devoid of pigmentation (albino). Populations can be limited to two discrete morphs, as seen in sexual dichotomy, or multiple morphs as seen with the harlequin ladybird (*Harmonia axyridis*), which can occur as a result of genetic variation or environmental pressures (White & Kemp 2016). Some eusocial insects also exhibit polymorphism, such as termites (Order: Blattodea), and ants, bees and wasps (Order: Hymenoptera), depending upon their function within their colony. For example, bees have three castes: queens – breeding females that are larger than the rest of the colony; workers – females that transport nectar and pollen using ‘pollen baskets’ on their hind legs; and drones – males whose function is to breed with a queen. These different roles result in morphological differences, and for some species all castes exhibit the same colours and patterns, while castes of other species differ in colouration. Sexual dimorphism is also evident across a variety of taxon groups. It can occur in mammals, birds, herpetofauna, invertebrates, and plants (Fairbairn & Blanckenhorn 2007), and variation between sexes can occur in form, colour, size and behaviour (Blanckenhorn 2005), but not always in the same direction (Owens & Hartley 1998). For example, some male spiders are a fraction of the size and different in colour to females (Vollrath & Parker 1992) (Fig. 1.2), while male southern sea lions (*Otaria byronia*) are larger than females from birth (Cappozzo & Campagna 1991). Such gender differences have resulted in at least one species being described as two separate species. One example is the eclectus parrot (*Eclectus roratus*). Vivid colouration in birds was associated only with males, and as the female was a vibrant red and blue colour (Fig. 1.2) it was assumed to be a different species altogether (Heinsohn 2005).

(a) news.nationalgeographic.com (b) touchn2btouched.tumblr.com

Figure 1.2 Examples of sexual dimorphism (a) Male (top) and female golden orb-weaver (genus *Nephila*) and (b) eclectus parrot (*Eclectus roratus*) female (in nest) and male.

There are also examples of species imitating other species. Evolution of mimicry in nature is widespread (Wallace 2010), with visual mimicry not only being found in a range of different taxon groups, but also linked to warning signals (aposematism). These signals involve both colour and striations, as animals with one type of photoreceptor (monochromats) only see in shades of grey. Different forms of mimicry include innocuous organisms mimicking the conspicuous markings of poisonous ones (Batesian) (Bates 1862), and mimicry between harmful species sharing a common predator (Müllerian) (Kapan 2001). Excellent identification skills may help to differentiate between mimics, but a prior knowledge of mimicry is likely to encourage an observer to explore certain features that may be otherwise overlooked.

Centuries of studying the natural world have provided a vast array of knowledge concerned with species, their habitats and ecosystems. However the *process* by which species identification occurs has not been widely studied. Furthermore, little is known about accuracy rates among individuals, the effects of expertise and experience, or whether differences in accuracy differ according to the morphology, life history or function of the organisms being identified. To explore these unknowns, a principled, simple method to assess identification accuracy is required, and one that does not rely on prior species knowledge. Furthermore, any investigation must be sensitive enough to capture individual differences.

One area of research that could help explore such patterns and processes is forensic face recognition. Understanding how humans identify, or do not identify, faces has been researched extensively over the past 30 years in the domain of Psychology (Young & Bruce 2011). Forensic face recognition shows that despite photographs being routinely used to establish people’s identity, visual identification can be surprisingly difficult (Bindemann & Sandford 2011). It also shows that individual ability to accurately identify faces is highly variable. This field of research not only has the potential to provide an understanding of how humans identify other species, but could also provide a model by which to test identification abilities. Applying these methods to species identification also has the potential to inform the development of training aids (e.g. identification guides and training courses), and ascertain if conservation decisions based on species data could be better informed.

1.3 Face recognition

The human face has a simple internal template, consisting of two eyes above a nose and a mouth, plus variation in external features such as hair, facial hair and eye colour (Longmore, Liu & Young 2014). The ability to easily recognise numerous different faces, even from poor quality images (Burton et al. 1999), led to the common assumption that humans are accurate in face identification (Hancock, Bruce & Burton 2000; Bahrick, Bahrick & Wittlinger 1975). In fact, while humans often believe themselves to be expert in recognising faces, problems in eye witness testimony show that we are good at identifying familiar faces, but surprisingly poor at recognising unfamiliar ones (Jenkins & Burton 2011).

The study of face recognition has generated extensive literature on identification and the impacts of *mis*identification. The identification of faces is routine in a number of everyday tasks, for example, in national security and crime investigation. Yet, despite its widespread use, it is known to be fallible and highly error-prone (Wells 1993; Jenkins & Burton 2008). Forensic face matching is the comparison of a pair of faces for identification purposes (Burton et al. 2010), which can be the same person (i.e. a ‘match’) or different people (i.e. a ‘mismatch’). Accuracy in identifying whether two images are of the same or different people has proved to be more difficult than expected (Kemp, Towell & Pike 1997), as there are a number of different limitations to photo identification (Megreya, Sandford & Burton 2013). For example, *different* photos of the *same* person are frequently thought to represent different people (Fig. 1.3) (Bindemann & Sandford 2011; Jenkins et al. 2011). Variation in illumination, distance, and different backgrounds can create environmental noise (Jenkins & Burton 2011; Lampinen et al. 2014), and the comparison of images differing in quality can reduce accuracy (Bruce et al. 2001). However, accuracy has been shown to improve when both images are of high quality (Henderson, Bruce & Burton 2001).

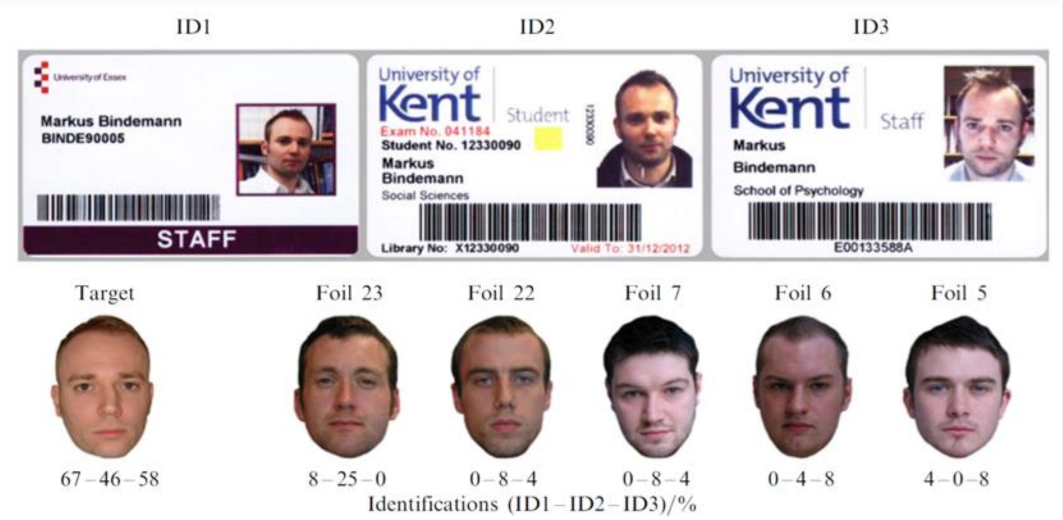


Figure 1.3. An illustration from Bindemann & Sandford (2011) showing the variation in photographs of the same individual taken at different times and varying conditions. Accuracy in matching the target to the images was 67% for ID1, 46% for ID2, and 58% for ID3, with 38% of observers (*n* = 24) accurately matching all three identification cards.

Studies in applied settings also show poor performance. In one study, over 50% of fraudulent photographic credit cards were falsely accepted by supermarket cashiers (Kemp, Towell & Pike 1997). Moreover, the cashiers were aware that they were participating in a study, plus the photographs were of high quality and taken no more than six weeks prior to the study. Another study found that on average, passport officers falsely accepted 14% of fraudulent photos, again based on high-quality photographs taken a few days before the task (White et al. 2014). Furthermore, individual officer performance was highly variable, independent of training or experience, and overall comparable with student participants at an average error rate of 20% (White et al. 2014). There is some evidence to suggest that experts trained in facial identification are consistently better at identifying CCTV images than members of the public (Wilkinson & Evans 2009). However, this study did not explore the identification accuracy of individuals, and some of the public participants may have performed at expert level.

Progress in face recognition has been limited by methodologies, and a lack of ecological validity (i.e. the delivery of an experiment in an attempt to closely emulate real-world conditions) (Burton 2013). Until recently, emphasis has been on matching *images* rather than matching *identities*. There has also been a focus on *between*-person variability (i.e. only using one image per person in face matching tasks). However, studies using different images of the same person demonstrate that *within*-person variability is considerable (Bindemann & Sandford 2011; Burton 2013). Many previous experiments have been delivered under idealised conditions, none of which are representative of real life situations, such as the use of a single image to represent one identity, using different images of individuals taken on the same day, or using images that have all been taken with the same image capture equipment, (Jenkins et al. 2011; Burton 2013; Papesh & Goldinger 2014). However, while a controlled laboratory environment is unlike everyday scenarios, too many distractors may add unnecessary environmental noise (Clutterbuck & Johnston 2005). By conducting research under ‘optimised conditions’, studies use stimuli that contain within person variation, as well as using different image capture equipment. However, as the tasks are researching *visual cues only*, the stimuli (images) lack context and personal information. For example, the target and foils (imposters) in

Fig 1.3 are presented as faces against a white background and all face forward, but they vary in illumination and camera angle, showing that no face presents the same image more than once (Jenkins & Burton 2011). However, there is one crucial variation that is participant dependent. When accuracy is investigated in terms of whether the participant is familiar or unfamiliar with the subject, results are remarkably different.

Matching familiar faces is performed with high precision (Jenkins et al. 2011), yet when the face is unfamiliar, the results can be very poor (Johnston & Edmonds 2009). Humans have the ability to recognise family and friends in images that vary greatly with pose, age, lighting etc., but the ability to do so with strangers is surprisingly error-prone. Bruce et al. (1999) found accuracy with unfamiliar targets to be 70%, even under optimal conditions, and Bindemann & Sandford (2011) reported accuracy ranging from 60% to 85%, despite the photographic identification presented being of the same person. Determining the point at which the unfamiliar become familiar is difficult to pinpoint, but the brain is known to process familiar and unfamiliar faces differently (Clutterbuck & Johnston 2005). Traditional research has focused on *recognising* familiar faces and *memorising* unfamiliar ones, but the importance of research in matching unfamiliar faces, especially given its widespread application in security settings, led to an increase in research in this topic. One of the first to do this was The Glasgow Face Matching Test (GMFT) (Burton, White & McNeill 2010), which is publically available and still used by researchers. The GMFT was a ‘new’ test for unfamiliar face matching using *different* images of the *same* person and found the mean accuracy to be 90%, ranging from 62% – 100%, with performance being better on matches (92%) than mismatches (88%) (Fig. 1.4).

However, it should be noted that for mismatches, images are selected to include two people that look similar. This may be simply the same gender, or could be the inclusion of similar external features (e.g. similar hairs colour, hair style, glasses).

Figure 1.4. Examples from The Glasgow Face Matching Test (A) is two different people (mismatch) and (B) are the same person (match). (From Burton et al. 2010).

Forensic face recognition provides a template for assessing visual identification accuracy. The application of these methods to species identification could provide a general indication of observer accuracy, explore differences between those familiar and unfamiliar with species, and investigate individual consistency. The findings could prove to be important to the management of biodiversity, as there is little evidence on how accurate species observation data is, or how it varies within and between observers. Matching tasks are relatively simple to deliver, and replicate the process by which an observer may compare an image of an unknown specimen with images of known specimens to decide if they are the same species. They are also independent of prior species knowledge, which is important for exploring the influence of experience and expertise on identification accuracy. Species data are collected by novice and experienced observers, although the difference between a beginner and an expert is not always clear or permanent.

1.4 Experts and expertise

What makes an expert? Expert knowledge could be regarded as knowledge on a subject not universally known (Martin et al. 2012), but describing an individual as an expert is subjective, possibly perceived differently by people outside a community than those within it, and can be controversial (Lave & Wenger, 1991; Goldman, 2001; Ericsson, 2014; Burgman, 2015). Society relies on individuals with relevant skills and experience for advice (Burgman, 2015), and while experts are a crucial resource in decision-making, their knowledge should be used appropriately (Stilgoe & Jones 2006; Sutherland & Burgman 2015). However, knowledge surrounding a subject can change, sometimes rapidly, and individuals within a domain also change (Fig. 1.5). Although potentially problematic to those seeking expert advice, disagreement between experts is to be expected, but overconfidence can result in expert judgement not reflecting the uncertainty involved (Martin et al. 2012; Dieckmann et al. 2017).

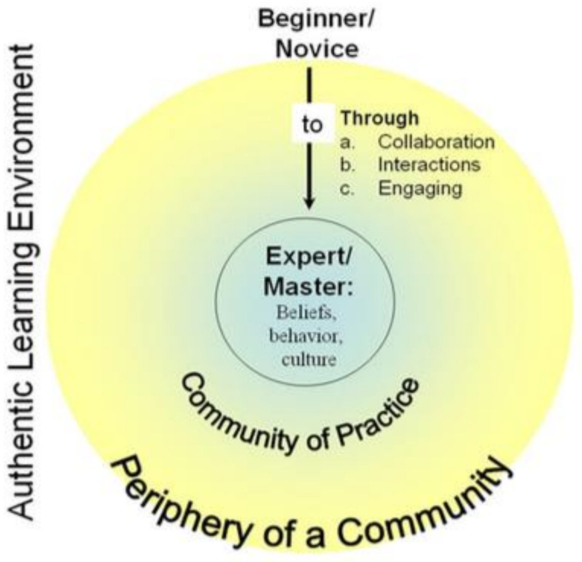


Figure 1.5. Model of a Community of Practice (illustration based on Lave & Wenger 2009). This figure illustrates the stages of member integration into a community of practice. Those at the centre may also leave the community.

Semantics around the use of ‘expert’ and ‘amateur’ can be problematic, especially given that ‘amateur’ can be used colloquially to describe incompetence rather than an unpaid or in a voluntary role. However, face recognition studies show that the ability to accurately match faces is not necessarily linked to training or experience. Burton et al. (1999) found that police officers with experience in forensic identification did not perform any better than students, which corresponds with White et al. (2014) findings that passport officers and students performed similarly (see section 1.3). However, there is an emergence of individuals with exceptional identification skills, known as “super-recognisers” (Russell, Duchaine & Nakayama 2009). These abilities may be evident to individuals before they seek employment in professions that perform face-matching tasks (e.g. police forces, passport officers, and security agencies), but can become apparent through internal recruitment or training (Bobak, Hancock & Bate 2016; Robertson et al. 2016). It could follow that species identification in could benefit from recruiting similar individuals from within and outside of the conservation community.

There is some debate over whether experienced taxonomists would be the only people with the ability to accurately identify certain species (Hopkins & Freckleton 2002). However, while experience can lead to a process becoming more automatic with a reduction in the effort required to complete a task, it is not necessarily linked to improved performance (Hoffman, Crandall & Shadbolt 1998; Ericsson 2014). Prior knowledge of a species can be advantageous, for example knowing what characteristics can aid accurate identification, the existence of various morphs, or known species distribution. There is an argument over whether taxonomists are declining in number (Hopkins & Freckleton 2002; Joppa, Roberts & Pimm 2011; Joppa, Roberts & Pimm 2012), but taxonomy complements other sources of information, for example field skills, rather than supersedes them. Identification skills are used both in the field and away from the field. Museum (Ballard et al. 2016) and biological (Consorte-McCrea et al. 2017) collections can contribute to species knowledge and often act as reference points in species identification. Some collections are accessible and some information is disseminated through websites, but there is no guarantee that the original identification is accurate (Goodwin et al. 2015).

Although visual identification skills are paramount in taxonomy, it does not follow that these skills are limited to the taxonomic community, or that accuracy is higher among professional taxonomists. Species observations are recorded in both professional and amateur capacities, and despite some criticism around the identification skills of amateur naturalists (Cohn 2008), they can be highly accurate (Ratnieks et al. 2016). These ‘amateurs’ have made a large contribution towards the knowledge of species numbers and distributions (Roy, Preston & Roy 2015), and the increasing contribution of volunteers is of value (Chandler et al. 2016). While ‘expert amateurs’, sometimes referred to as citizen scientists, range in their skills, education, culture and experience, they share an interest which continues to contribute to our knowledge of the natural world.

‘Citizen science’ is a relatively new term for an old practice. The term ‘scientist’ was only coined in the 19th Century (previously ‘natural philosophers’) and the concept of a professional scientist, especially in ecology, is a relatively recent one. Historically, many scientists were independently wealthy or made their income by other means (Silvertown 2009). The UK has a rich history of amateur naturalists (Allen 1976), with a wealth of scientific information coming from individuals who collected and studied specimens from across the world, many of which continue to be used in research today (Suarez & Tsutsui 2004). Contemporary pressures on the natural world have resulted in changes to the way species are studied, especially in terms of collecting. However, modern technology has created opportunities for data collection from previously inaccessible areas (Pimm et al. 2015) and enabled wider participation in these studies (e.g. Swanson et al. 2015). Citizen science is multi-faceted, lacks a single definition (Kullenberg & Kasperowski 2016), and is often a term used by academics rather than practitioners, but it is increasingly an umbrella term used for volunteer involvement.

Contributions from non-scientific participants to the study of the natural world is recognised in a number of environmental disciplines, including climate change, invasive species, conservation biology, and monitoring (Silvertown 2009). In general, projects using the term ‘citizen science’ engage volunteers in the collection, collation, processing and analysis of data, and can be applied to those practicing science in a volunteer or self-organised capacity. Citizen science can be ‘top-down’ or ‘bottom-up’, structured or unstructured, and systematic or ad-hoc. As data collection on a large scale is inherently expensive (Jones 2011), there is a growing reliance on the volumes of citizen science data for monitoring biodiversity (Ruiz-Gutierrez, Hooten & Grant 2016), which can be collected via targeted projects or downloaded from accessible datasets (see section 1.5.1). It is worth noting that these large volumes of data may not represent the entire picture, as there are always data waiting to be processed or digitised, for example natural history collections, paper records belonging to individuals, and observations submitted to conservation groups.

An element of distrust between science and volunteer communities has been alluded to, with some academics questioning the value of data collected by the public (Cohn 2008). Data are collected according to the needs of the collector and may not be suitable to those answering different questions or with different objectives. However, there is evidence of professional and amateur datasets being combined (Flesch & Belt 2017). Irrelevant of the capacity in which species observations are collected, little is known about how accurate these identifications are (Beerkircher, Arocha & Barse 2009).

1.5 Error rates

Species observations are known to contain error and bias. These include false positives, non-detection errors (Farmer, Leonard & Horn 2012), spatial bias (Geldmann et al. 2016) and recorder effort (Isaac & Pocock 2015). However, despite concerns relating to declines in identification skills, there are few studies that test species identification accuracy (Tillett et al. 2012). The studies that have investigated ability in species identification found that accuracy was variable between taxa and participants, and was influenced by participant interaction with the study species. For example, a study of dinoflagellate classification found identification accuracy among experts to be 72%, ranging from 38% to 95%, depending upon the species being identified (Culverhouse et al. 2003). In addition, experts with field expertise averaged 97% accuracy and were highly consistent, while those whose expertise came from books averaged 75% with more variable consistency (Culverhouse et al. 2003). This example found that over one-in-four identifications was a *mis*identification, and that although participants were experienced, how their experience was gained affected their decision-making. Species misidentification has been found to affect sustainable fishing, as confusion between some fish species (e.g. *Tetrapturus albidus* and *T. georgii*) results in false population assessments and possible mismanagement of fish stocks (Beerkircher, Arocha & Barse 2009; Kitchen-Wheeler 2010; Tillett et al. 2012), as well as the monitoring of invasive species (Robinson, Inger & Gaston 2016). There have also been studies into identifying individuals from images, for example mountain bongo antelopes (*Tragelaphus eurycerus isaaci*) which found errors in at least one-in-five trials (Gibbon, Bindemann & Roberts 2015), and Andean bears (*Tremarctos ornatus*) in which neither experience nor confidence were indicators of accuracy (Horn et al.

2014). These show that errors are made in the identification of both species and individuals, and while the judgement of what level of accuracy is considered ‘good’ or ‘poor’ is subjective, small percentage error rates may potentially have a large impact.

1.5.1 A question of scale

Forensic face matching studies have shown that even in comparatively controlled settings, accurate identification is far from perfect and even small error rates would translate into thousands of errors in large scale systems (Jenkins & Burton 2008b). For example, while there is no way of knowing how many identity checks occur on a daily basis, the figures for air travel alone are substantial. INTERPOL note that in addition to the 800 million searches of their database to screen passports in 2013, there were an estimated 1 billion occasions where passenger passports went unchecked on international flights in the same period (INTERPOL 2014). For 2017, it is estimated that 1.5 billion passengers will take international flights (INTERPOL 2014). Although it is hard to estimate how many species observations exist, biological collections, such as museums, hold millions of observations in the form of specimens. There are also some large datasets that give some idea of the volume of data being stored and widely used on a regular basis.

Species observations are used in a variety of ways and provide important information to individuals and organisations. These can be from organised events such as BioBlitz (www.bnhc.org.uk) where people aim to record as many species as they can in one area in a set amount of time, ad-hoc observations, or systematically collected by amateurs and researchers alike. Some examples of large datasets and their uses are:

- **eBird** started in 2002, this project comprises an online checklist of global bird species. Observations are submitted by worldwide ‘recreational and professional’ birdwatchers, and these openly available data have a wide spectrum of end users, including government, policy-makers, land managers, researchers and NGOs. The dataset is vast, with 9.5 million observation being submitted in May 2015 alone (Sullivan et al. 2014);
- **Global Biodiversity Information Facility (GBIF)** was formed to share biodiversity information, advance scientific research and avoid duplication of effort. Established in 2001 by the Organization for Economic Cooperation and Development (OECD), it holds data on 1.6 million species and is funded by a number of governments (GBIF 2017);
- **The International Union for Conservation of Nature (IUCN)** is an environmental network and considers itself “the global authority on the status of the natural world and the measures needed to safeguard it”. It has approximately 1,300 member organisations and since the 1950s has produced lists of threatened species in its ‘Red Lists’ categorising their risk of extinction (Butchart et al. 2005). These data are biased towards terrestrial megaflora and megafauna in temperate environments (Baillie, Collen & Amin 2008), but in some cases may be the only data available (Molinari- Jobin et al. 2012);
- **The NBN** is a UK repository and holds over 132 million records (National Biodiversity Network 2017). It is a charity through which government agencies, NGOs, volunteer groups and individuals exchange biodiversity information.

In these few examples alone, there are hundreds of millions of records that are freely available and used in numerous decision-making processes. As with many processes, there is a possibility of error, and with such vast numbers of observations even small error rates could have an impact. However, whether errors exist and where they might occur is rarely questioned, and this thesis aims to explore this.

1.6 Outline of this thesis

This thesis applies methods established in forensic face matching research to investigate accuracy in species identification. There is a growing demand for data to provide evidence on species numbers, diversity, distribution, and interactions, yet associated literature rarely discusses testing the identification ability of data collectors, or possible errors that could occur in those data. Advances in technology mean that previous restrictions to data collection (e.g. lack of access due to location or activity window) have diminished thanks to the capability of image capture equipment (e.g. photographs and videos from smart phones, camera traps and drones). In addition, the falling cost of certain makes and models has made these tools widely available, and images form an important part of data collection for both professional and lay scientists, which in turn are easily shared electronically.

The specific aims of this thesis are:

- To investigate whether familiarity with a taxon group increases accuracy in species identification;
- To explore whether differences in levels of expertise correlates with observer ability to accurately recognise images as the same or different species;
- To examine levels of consensus within observers of the same level of expertise;

- To explore consistency within and between observers;
- To investigate the effect of intra-specific variation on choices made due to inter-specific variation.

Chapter 2 acts as a starting point to investigate the effects of familiarity on accurate species identification. Pollinators are of great conservation concern, and by selecting native bumblebees (*Bombus*) as model species, unfamiliar participants are likely to have encountered a bumblebee, but will not necessarily be aware of how many bumblebee species occur in the UK. A repeated, pairwise matching task is delivered online, and allows direct comparison of observers familiar with the model species to those without prior knowledge. The images in this task are taken from popular identification guides and presented under optimised conditions, thus limiting decisions on images being the same species to visual cues alone. This task will explore variation both among individuals and within expertise groups, plus consistency in identification decisions.

Chapter 3 builds on the work of **Chapter 2**. Here, only individuals unfamiliar with bumblebees participate in the study. A pairwise matching task compare queens of 20 different bumblebee (*Bombus*) species with each other. Once again, the task is delivered online with images presented under optimised conditions, but is not repeated. The task tests variation of accuracy among individuals with little or no experience with these species, and investigates whether inter-species within the same genus is perceived in a uniform manner.

Chapter 4 investigates judgement and consensus of experts with a simple sorting task. Conservation practitioners in possession of the relevant licence to survey for great crested newts (*Triturus cristatus*) are given 80 newt images to sort into species. The images comprise three native and one non-native newt, all of which may be encountered when surveying in the UK. Participants individually group images by species, and then name that species. This task tests the identification abilities of experts who have achieved the same level of proficiency within a knowledgeable community. It also reports individual performance in relation to self-perceived abilities and experience, plus the extent to which participants agreed on identification.

Chapter 5 investigates variation in choice by novices when sorting images into species groups. With the same stimuli images as used in **Chapter 4**, participants are asked to group images by species, initially without knowing how many species there are, and then again after being informed that four species are present. This task explores judgements, both within and between non-experts participants, on ambient images that are presented simultaneously. Unlike the images presented to non-expert participants under optimised conditions in **Chapter 2**, these images include context, and reflect the intra- and inter-species variation encountered in the field. The task also determines if any species are grouped together more frequently than others.

1.7 References

Agapow, P.M., Bininda-Emonds, O.R., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C. and Purvis, A. (2004). The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*, **79**, 161–179.

Allen, D. (1976). *The Naturalist in Britain: A Social History*. Princeton University Press, Princeton.
https://askabiologist.asu.edu/sites/default/files/resources/articles/monarchs/complete-metamorphosis.gif

Bahrick, H.P., Bahrick, P.O. and Wittlinger, R.P. (1975). Fifty years of memory for names and faces: a cross-sectional approach. *Journal of Experimental Psychology: General*, **104**, 54–75.

Baillie, J., Collen, B., Amin, R. Akcakaya, H.R., Butchart, S.H., Brummitt, N., Meagher, T.R., Ram, M., Hilton- Taylor, C. and Mace, G.M (2008). Toward monitoring global biodiversity. *Conservation Letters*, **1**, 18–26.

Ballard, H.L., Robinson, L.D., Young, A.N., Pauly, G.B., Higgins, L.M., Johnson, R.F. and Tweddle, J.C. (2017). Contributions to conservation outcomes by natural history museum-led citizen science: examining evidence and next steps. *Biological Conservation*, **208**, 87-97.

Bates, H.W. (1862). XXXII. Contributions to an insect fauna of the Amazon Valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London*, **23**, 495–566.

Beerkircher, L., Arocha, F., Barse, A., Prince, E., Restrepo, V., Serafy, J. and Shivji, M. (2009). Effects of species misidentification on population assessment of overfished white marlin *Tetrapturus albidus* and roundscale spearfish *T. georgii*. *Endangered Species Research,* **9**, 81-90.

Bindemann, M., Avetisyan, M. and Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, **18**, 277.

Bindemann, M. and Sandford, A. (2011). Me, myself, and I: different recognition rates for three photo-IDs of the same person. *Perception*, **40**, 625-627.

BirdLife International (2016). *Porphyrio hochstetteri*. The IUCN Red List of Threatened Species 2016: e.T22692808A93370351. Available from: http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22692808A93370351.en. [Accessed 29 December 2017].

Blanckenhorn, W. (2005). Behavioral causes and consequences of sexual size dimorphism. *Ethology,* **111**, 977-1016.

Bland, L. and Collen, B. (2016). Species loss: lack of data leaves a gap. *Nature,* **537**, 488-488.

Boakes, E., Rout, T. and Collen, B. (2015). Inferring species extinction: the use of sighting records. *Methods in Ecology and Evolution*, **6**, 678-687.

Boakes, E.H., McGowan, P.J., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. and Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology*, **8**, p.e1000385.

Bobak, A.K., Hancock, P.J.B. and Bate, S. (2016). Super-recognisers in action: evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, **30**, 81–91.

Briggs, S. and Hudson, M.D. (2013). Determination of significance in Ecological Impact Assessment: past change, current practice and future improvements. *Environmental Impact Assessment Review*, **38**, 16–25.

Bruce, V., Henderson, Z., Newman, C. and Burton, A.M (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, **7**, 207-218.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A.M. and Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, **5**, 339–360.

Burton, A.M., Wilson, S., Cowan, M. and Bruce, V. (1999). Face recognition in poor-quality video: evidence from security surveillance. *Psychological Science*, **10**, 243–248.

Burton, A.M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology* **66**, 1467-1485.

Burton, A.M., White, D. and McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, **42**, 286-291.

Butchart, S.H.M., Stattersfield, A.J., Baillie, J., Bennun, L.A., Stuart, S.N., Akçakaya, H.R., Hilton-Taylor, C. and Mace, G.M. (2005). Using Red List Indices to measure progress towards the 2010 target and beyond. *Philosophical Transactions of the Royal Society B*, **360**, 255–268.

Cappozzo, H., Campagna, C. and Monserrat, J. (1991). Sexual dimorphism in newborn southern sea lions. *Marine Mammal*, **7**, 385-394.

Chandler, M., Rullman, S., Cousins, J., Esmail, N., Begin, E., Venicx, G., Eisenberg, C. and Studer, M. (2016). Contributions to publications and management plans from 7 years of citizen science: use of a novel evaluation tool on Earthwatch-supported projects. *Biological Conservation*, **208**, 163-173.

Convention on International Trade in Endangered Species (*CITES*) [Online] (2017). *What is CITES?* Available from: https://cites.org/eng/disc/what.php [Accessed 1 May 2017].

Clutterbuck, R. and Johnston, R.A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, **17**, 97-116.

Cohn, J.P. (2008). Citizen Science: can volunteers do real research? *BioScience*, **58**, 192-197.

Consorte-McCrea, A., Bainbridge, A., Fernandez, A., Nigbur, D., McDonnell, S., Morin, A. and Grente, O. (2017). Understanding attitudes towards native wildlife and biodiversity in the UK: the role of zoos. In: Filho, W.L. ed. *Sustainable Development Research at Universities in the United Kingdom*. Springer, London, pp. 295-311.

Crous, P.W., Groenewald, J.Z., Slippers, B. and Wingfield, M.J. (2016). Global food and fibre security threatened by current inefficiencies in fungal identification. *Philosophical Transactions of the Royal Society of London B*, **371**, p.20160024.

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, **247**, 17–25.

Darwin, C. (1859). *The origin of species*. Gramercy Books, New York.

Diaz, J.H. (2016). Poisoning by herbs and plants: rapid toxidromic classification and diagnosis. *Wilderness & Environmental Medicine*, **27**, 136-152.

Díaz, S., Fargione, J., Chapin III, F.S. and Tilman, D. (2006). Biodiversity loss threatens human well-being. *PLoS Biology*, **4**, e277.

Dieckmann, N.F., Johnson, B.B., Gregory, R., Mayorga, M., Han, P.K. and Slovic, P. (2017). Public perceptions of expert disagreement: bias and incompetence or a complex and random world? *Public Understanding of Science*, **26**, 325–338.

Donaldson J and Scheffler, A. (1999). *The Gruffalo*. Macmillan Publishers, London.

Elphick, C.S. (2008). How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, **45**, 1313–1320.

Ericsson, K.A. (2014). Expertise. *Current Biology*, **24**, R508–R510.

Evans, M., Merow, C., Record, S., McMahon, S.M. and Enquist, B.J. (2016). Towards process-based range modeling of many species. *Trends in Ecology & Evolution*, **31**, 860-871.

Fairbairn, D.J., Blanckenhorn, W. and Székely, T. (2007). *Sex, size, and gender roles: evolutionary studies of sexual size dimorphism.* Oxford University Press, Oxford.

Farhan Ferrari, M., de Jong, C. and Belohrad, V.S. (2015). Community-based monitoring and information systems (CBMIS) in the context of the Convention on Biological Diversity (CBD). *Biodiversity*, **16**, 57-67.

Farmer, R., Leonard, M. and Horn, A. (2012). Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. *The Auk* **129**, 76-86.

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J. Stevenson, R.D., Courtney, G.W., Van Dyk, J.K. and Ellison A.M. (2013). Next-generation field guides. *BioScience,* **63**, 891–899.

Flesch, E.P. and Belt, J.J. (2017). Comparing citizen science and professional data to evaluate extrapolated mountain goat distribution models. *Ecosphere*, **8**, e01638.

Forêt, S., Knack, B., Houliston, E., Momose, T., Manuel, M., Quéinnec, E., Hayward, D.C., Ball, E.E. and Miller, D.J. (2010). New tricks with old genes: the genetic bases of novel cnidarian traits. *Trends in genetics,* **26**, 154–158.

Freudenstein, J.V., Broe, M.B., Folk, R.A. and Sinn, B.T. (2016). Biodiversity and the species concept—lineages are not enough. *Systematic Biology*, **66**, 644-656.

Gallai, N., Salles, J-M., Settele, J. and Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, **68**, 810–821.

Gaston, K.J. (2000). Global patterns in biodiversity. *Nature* **405**, 220–227.

Geldmann, J., Heilmann-Clausen, J., Holm, T.E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C. and Tøttrup, A.P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, **22**, 1139–1149.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ*, **3**, e1303.

Global Biodiversity Information Facility (GBIF) (2017) *Free and open access to biodiversity data* [Online]. Available from: http://www.gbif.org/ [Accessed: 20 February 2017].

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I. and Scotland, R.W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, **25**, R1066–R1067.

Government of Western Australia, Department of Agriculture and Food (2017). *MyPestGuide* [Online]. Available from: https://agspsrap31.agric.wa.gov.au/mypestguide/#/ [Accessed 22 June 2017].

Hancock, P.J.B., Bruce, V. and Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, **4**, 330–337.

Hardy, M. and Barrington, D. (2017). A transdisciplinary approach to managing emerging and resurging mosquito-borne diseases in the Western Pacific Region. *Tropical Medicine and Infectious Disease*, **2**, 1.

Harlequin Ladybird Survey, The (2017). *Recording sightings* [Online]. Available from: http://www.harlequin-survey.org/recording.htm. [Accessed 22 June 2017].

Harrison, J.R., Roberts, D.L. and Hernandez-Castro, J. (2016). Assessing the extent and nature of wildlife trade on the dark web. *Conservation Biology*, **30**, 900–904.

Heinsohn, R., Legge, S. and Endler, J.A. (2005). Extreme reversed sexual dichromatism in a bird without sex role reversal. *Science*, **309**, 617-619.

Henderson, Z., Bruce, V. and Burton, A.M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, **15**, 445–464.

Hendry, A.P., Vamosi, S.M., Latham, S.J., Heilbuth, J.C. and Day, T. (2000). Questioning species realities. *Conservation Genetics*, **1**, 67–76.

Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773-785.

Hey, J. (2006). On the failure of modern species concepts. *Trends in Ecology & Evolution*, **21**, 447–450.

Hey, J. (2001). The mind of the species problem. *Trends in Ecology & Evolution*, **16**, 326–329.

Heywood, V.H. (1998). The species concept as a socio-cultural phenomenon – a source of the scientific dilemma. *Theory in Biosciences*, **117**, 203–212.

Hill, A., Guralnick, R., Smith, A., Sallans, A., Gillespie, R., Denslow, M., Gross, J., Murrell, Z., Conyers, T., Oboyski, P., Ball, J., Thomer, A., Prys-Jones, R., de la Torre, J., Kociolek, P., Fortson, L., (2012). The notes from nature tool for unlocking biodiversity records from museum records through citizen science. *ZooKeys*, **209**, 219–233.

Hinder, S., Hays, G.C., Edwards, M., Roberts, E.C., Walne, A.W. and Gravenor, M.B. (2012). Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*, **2**, 271-275.

Hinsley, A., King, E. and Sinovas, P. (2016). Tackling illegal wildlife trade by improving traceability: a case study of the potential for stable isotope analysis. In: Potter, G.R., Nurse, A. and Hall, M. eds. *The Geography of Environmental Crime*. Palgrave Macmillan UK, London. 91–119.

Hoffman, R.R., Crandall, B. and Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: a case study in the methodology of cognitive task analysis. *Human Factors*, **40**, 254–276.

Hopkins, G.W. and Freckleton, R.P. (2002). Declines in the numbers of amateur and professional taxonomists : implications for conservation. *Animal Conservation*, **5**, 245–249.

Hoser, R.T. (2015). The Wüster gang and their proposed “Taxon Filter”: how they are knowingly publishing false information, recklessly engaging in taxonomic vandalism and directly attacking the rules and stability of zoological nomenclature. *Australasian Journal of Herpetology*, **25**, 14–38.

Humphries, C.J., Williams, P.H. and Vane-Wright, R.I. (1995). Measuring biodiversity value for conservation. *Annual review of ecology and systematics*, **26**, 93-111.

Hunt, E. (2015). *New Zealand hunters apologise over accidental shooting of takahē*. The Guardian [Online], (Environment) 21 August. Available from: http://www.theguardian.com/environment/2015/aug/21/new-zealand-conservationists-apologise-over-accidental-shooting-of-endangered-takahe [Accessed 21st August 2015].

Isaac, N.J.B. and Pocock, M.J.O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, **115**, 522-531.

INTERPOL (2014). *INTERPOL confirms at least two stolen passports used by passengers on missing Malaysian Airlines flight 370 were registered in its databases.* Available from: https://www.interpol.int/News-and-media/News/2014/N2014-038 [Accessed: 16 May 2017].

Jenkins, R. and Burton, A. (2008). Limitations in facial identification: The evidence. *Justice of the Peace*, **172**, 4-6.

Jenkins, R. and Burton, A.M. (2008b). 100% accuracy in automatic face recognition. *Science*, **319**, 435.

Jenkins, R. and Burton, A.M. (2011). Stable face representations. *Philosophical transactions of the Royal Society B*, **366**, 1671-1683.

Jenkins, R. White, D., Van Montfort, X. and Burton, A.M. (2011). Variability in photos of the same face. *Cognition*, **121**, 313-323.

Johnston, R.A. and Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: a review. *Memory,* **17**, 577-596.

Jones, J.P.G. (2011). Monitoring species abundance and distribution at the landscape scale. *Journal of Applied Ecology*, **48**, 9–13.

Joppa, L.N., Roberts, D.L. and Pimm, S.L. (2012). Taxonomy that matters: response to Bacher. *Trends in Ecology and Evolution*, **27**, 66.

Joppa, L.N., Roberts, D.L. and Pimm, S.L. (2011). The population ecology and social behaviour of taxonomists. *Trends in Ecology and Evolution*, **26**, 551–553.

Kapan, D.D. (2001). Three-butterfly system provides a field test of Müllerian mimicry. *Nature*, **409**, 338–340.

Kassin, S.M., Dror, I.E. and Kukucka, J. (2013). The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, **2**, 42–52.

Kemp, R., Towell, N. and Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud. *Applied Cognitive Psychology*, **11**, 211–222.

Kitchen-Wheeler, A-M. (2010). Visual identification of individual manta ray (*Manta alfredi*) in the Maldives Islands, Western Indian Ocean. *Marine Biology Research*, **6**, 351–363.

Krell, F.T. (2004). Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of ‘morphospecies’ sorting. *Biodiversity and Conservation*, **13**, 795–812.

Kullenberg, C. and Kasperowski, D. (2016). What is citizen science? - a scientometric meta-analysis. *PLoS One*, **11**, e0147152.

Kunz, W. (2012). *Do Species Exist? Principles of Taxonomic Classification*. Wiley-VCH Verlag & Co., Weinheim.

Lampinen, J.M., Erickson, W.B., Moore, K.N. and Hittson, A. (2014). Effects of distance on face recognition: implications for eyewitness identification. *Psychonomic bulletin & review*, **21**, 1489–1494.

Latombe, G., Pyšek, P., Jeschke, J.M., Blackburn, T.M., Bacher, S., Capinha, S.B., Costello, M.J., Fernández, M., Gregory, R.D., Hobern, D., Hui, C., Jetz, W., Kumschick, S., McGrannachan, C., Pergl, J., Roy, H.E., Scalera, R., Squires, Z.E., Wilson, J.R.U., Winter, M. Genovesi, P. and McGeoch, M.A. (2016). A vision for global monitoring of biological invasions. *Biological Conservation* (in press).

Lewis, B., Griffiths, R.A. and Wilkinson, J.W. (2017). Population status of great crested newts (*Triturus cristatus*) at sites subjected to development mitigation. *Herpetological Journal*, **27**, 133–142.

Lindenmayer, D.B. and Likens, G.E. (2010). The science and application of ecological monitoring. *Biological Conservation*, **143**, 1317–1328.

Locey, K.J. and Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 201521291.

Longmore, C.A., Liu, C.H. and Young, A.W. (2014). The importance of internal facial features in learning new faces. *The Quarterly Journal of Experimental Psychology*, **68**, 249-260.

Mallet, J. (2001). Species, concepts of. *Encyclopedia of biodiversity*, **5**, 427-440.

Martin, T., Burgman, M., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. and Mengersen, K. (2012). Eliciting expert knowledge in conservation science. *Conservation*, **26**, 29-38.

Maxwell, S., Fuller, R.A., Brooks, T.M. and Watson, J.E. (2016). Biodiversity: the ravages of guns, nets and bulldozers. *Nature*, **536**, 143-145.

Mayden, R.L. (2002). On biological species, species concepts and individuation in the natural world. *Fish and Fisheries*, **3**, 171–196.

Mayr, E. (1942). *Systematics and the Origin of Species, from the viewpoint of a zoologist*. Harvard University Press.

Megreya, A.M., Sandford, A. and Burton, A.M. (2013). Matching face images taken on the same day or months apart: the limitations of photo ID. *Applied Cognitive Psychology*, **27**, 700-706.

Memmott, J., Waser, N.M. and Price, M. V. (2004). Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society B,* **271**, 2605-2611.

Molinari-Jobin, A., Kéry, M., Marboutin, E., Molinari, P., Koren, I., Fuxjäger, C., Breitenmoser- Würsten, C., Wölfl, S., Fasel, M., Kos, I. and Wölfl, M. (2012). Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. *Animal Conservation*, **15**, 266-273.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G. and Worm, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biology*, **9**, e1001127.

Mysterud, A., Easterday, W.R., Stigum, V.M., Aas, A.B., Meisingset, E.L. and Viljugrein, H. (2016). Contrasting emergence of Lyme disease across ecosystems. *Nature Communications*, **7**, 11882.

National Biodiversity Network (NBN) (2017) [Online]. Available from: https://nbn.org.uk/ [Accessed: 21 February 2017].

Nellemann, C., Henriksen, R., Kreilhuber, A., Stewart, D., Kotsovou, M., Raxter, P., Mrema, E. and Barrat, S. eds. (2016). *The rise of environmental crime: a growing threat to natural resources, peace, development and security.* United Nations Environment Programme (UNEP), Nairobi.

Office for National Statistics (ONS) (2017). *Construction Statistics* [Online]. Available from: https://www.ons.gov.uk/businessindustryandtrade/constructionindustry/articles/constructionstatistics/no172016edition [Accessed: 12 May 2017].

Oldfield, S. ed. (2003). *The Trade in Wildlife: regulation for conservation*. Earthscan Publications Ltd., London.

Owens, I.P.F. and Hartley, I.R. (1998). Sexual dimorphism in birds: why are there so many different forms of dimorphism? *Proceedings of the Royal Society* B, **265**, 397-407.

Papesh, M.H. and Goldinger, S.D. (2014). Infrequent identity mismatches are frequently undetected. *Attention, Perception & Psychophysics* **76**, 1335–1349.

Phelps, J. and Webb, E.L. (2015). ‘Invisible’ wildlife trades: Southeast Asia’s undocumented illegal trade in wild ornamental plants. *Biological Conservation* **186**, 296–305.

Pimm, S., Russell, G.J., Gittleman, J.L. and Brooks, T.M. (1995). The future of biodiversity. *Science,* **296**, 347.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, **30**, 685-696.

Powney, G.D. and Isaac, N.J.B. (2015). Beyond maps: a review of the applications of biological records. *Biological Journal of the Linnean Society*, **115**, 532–542.

Purvis, A., Gittleman, J.L., Cowlishaw, G. and Mace, G.M., (2000). Predicting extinction risk in declining species. *Proceedings The Royal Society B,* **267**, 1947–1952.

Pyle, R.L. and Michel, E. (2008). ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. *Zootaxa*, **1950**, 39–50.

De Queiroz, K. (2005). Colloquium Paper: Systematics and the Origin of Species: Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, **102**, 6600-6607.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, **56**, 879–886.

Rare Newts 'Planted' at Development Sites to Stop the Bulldozers (2016). *Telegraph* [Online], 12 October. Available from: http://www.telegraph.co.uk [Accessed: 12 May 2017].

Ratnieks, F.L.W., Schrell, F., Sheppard, R.C., Brown, E., Bristow, O.E. and Garbuzov, M. (2016). Data reliability in citizen science: learning curve and the effects of training method, volunteer background and experience on identification accuracy of insects visiting ivy flowers. *Methods in Ecology and Evolution*, **7**, 1226-1235.

Relyea, R.A. (2004). Fine-tuned phenotypes, tadpole plasticity under 16 combinations of predator and competitors. *Ecology*, **85**, 172-179.

Renner, S.S. (2016). A return to Linnaeus’s focus on diagnosis, not description: the use of DNA characters in the formal naming of species. *Systematic Biology*, **65**, 1085–1095.

Roberts, D.L., Elphick, C.S. and Reed, J.M. (2010). Identifying anomalous reports of putatively extinct species and why it matters. *Conservation biology*, **24**, 189–196.

Robertson, D.J., Noyes, E., Dowsett, A.J., Jenkins, R. and Burton, A.M. (2016). Face recognition by Metropolitan Police super-recognisers. *PLoS One*, **11**, e0150036.

Robinson, B.S., Inger, R. and Gaston, K.J. (2016). A rose by any other name: plant identification knowledge & socio-demographics. *PLoS One*, **11**, e0156572.

Rodrigues, A.S.L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M. and Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in Ecology and Evolution*, **21**, 71–76.

Roy, H.E. and Brown, P.M.J. (2015). Ten years of invasion: *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae) in Britain. *Ecological Entomology*, **40**, 336-348.

Roy, H.E., Preston, C.D. and Roy, D.B. (2015). Fifty years of the Biological Records Centre. *Biological Journal of the Linnean Society*, **115**, 469–474.

Ruiz-Gutierrez, V., Hooten, M.B. and Campbell Grant, E.H. (2016). Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, **7**, 900-909.

Russell, R., Duchaine, B. and Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, **16**, 252–257.

Scott, P. and Rines, R. (1975). Naming the Loch Ness monster. *Nature*, **258**, 466–468.

Scriven, J.J., Woodall, L. C., Tinsley, M. C., Knight, M. E., Williams, P. H., Carolan, J. C., Brown, M. J. F., Goulson, D. (2015). Revealing the hidden niches of cryptic bumblebees in Great Britain: implications for conservation. *Biological Conservation*, **182**, 126–133.

Sih, A. and Christensen, B. (2001). Optimal diet theory: when does it work, and when and why does it fail? *Animal Behaviour*, **61**, 379–390.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, **24**, 467–471.

Solow, A., Smith, W., Burgman, M., Rout, T., Wintle, B. and Roberts, D. (2011). Uncertain sightings and the extinction of the ivory-billed woodpecker. *Conservation Biology*, **26**, 180–184.

Spamer, E.E. (1999). Know thyself: responsible science and the lectotype of *Homo sapiens* Linnaeus. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **149**, 109-114.

Stilgoe, J., Irwin, A. and Jones, K. (2006). The challenge is to embrace different forms of expertise to view them as a resource rather than a burden. *Opening up expert advice.* Demos, London.

Suarez, A. V. and Tsutsui, N.D. (2004). The value of museum collections for research and society. *BioScience,* **54**, 66-74.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A. and Fink, D. (2014). The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation,* **169**, 31-40.

Sutherland, W.J. and Burgman, M. (2015). Policy advice: use experts wisely. *Nature,* **526**, 317–318.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. and Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, **2**, 150026.

Tillett, B.J., Field, I.C., Bradshaw, C.J., Johnson, G., Buckworth, R.C., Meekan, M.G. and Ovenden, J.R. (2012). Accuracy of species identification by fisheries observers in a north Australian shark fishery. *Fisheries Research*, **127**, 109–115.

Treweek, J. (2009). *Ecological Impact Assessment.* John Wiley & Sons. London.

Turnhout, E., Lawrence, A. and Turnhout, S. (2016). Citizen science networks in natural history and the collective validation of biodiversity data. *Conservation Biology,* **30**, 532-539.

Van Horn, R.C., Zug, B., LaCombe, C., Velez-Liendo, X. and Paisley, S. (2014). Human visual identification of individual Andean bears *Tremarctos ornatus*. *Wildlife Biology*, **20**, 291–299.

Vane-Wright, R.I. (2000). Taxonomy, methods of. *Encyclopedia of Biodiversity,* **7**, 97–111.

Vaux, A., Gibson, G. Hernandez-Triana, L.M., Cheke, R.A., McCracken, F., Jeffries, C.L., Horton, D.L., Springate, S., Johnson, N., Fooks, A.R. and Leach, S. (2015). Enhanced West Nile virus surveillance in the North Kent marshes, UK. *Parasites & Vectors*, **8**, 91.

Vollrath, F. and Parker, G.A. (1992). Sexual dimorphism and distorted sex ratios in spiders. *Nature,* **360**, 156–159.

Wallace, A.R. (2016). *Mimicry, and Other Protective Resemblances Among Animals* (1867). Read Books Ltd., London.

Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist*, **48**, 553–571.

White, D., Kemp, R.I., Jenkins, R., Matheson, M. and Burton, A.M. (2014). Passport officers' errors in face matching. *PLoS One*, **9**, e103510.

White, T.E. and Kemp, D.J. (2016). Colour polymorphism. *Current Biology*, **26**, R517–R518.

Wilkins, J.S. (2011). Philosophically speaking, how many species concepts are there? *Zootaxa*, **2765**, 58-60.

Wilkinson, C. and Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science & Justice*, **49**, 191–196.

Will, K.W. and Rubinoff, D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, **20**, 47–55.

Williams, P.H., Brown, M.J., Carolan, J.C., An, J., Goulson, D., Aytekin, A.M., Best, L.R., Byvaltsev, A.M., Cederberg, B., Dawson, R. and Huang, J. (2012). Unveiling cryptic species of the bumblebee subgenus *Bombus s. str.* worldwide with COI barcodes (Hymenoptera: Apidae). *Systematics and Biodiversity*, **10**, 21–56.

Young, A.W. and Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, **102**, 959-974.

Parrot or guacamole? ©teenybiscuit

Chapter 2

Scientific Reports **6**, 33634 (2016).

2 Species identification by experts and non-experts: comparing images from field guides.

Gail E. Austen, Markus Bindemann, Richard A. Griffiths and David L. Roberts

2.1 Abstract

Accurate species identification is fundamental when recording ecological data. However, the ability to correctly identify organisms visually is rarely questioned. We investigated how experts and non-experts compared in the identification of bumblebees, a group of insects of considerable conservation concern. Experts and non-experts were asked whether two concurrent bumblebee images depicted the same or two different species. Overall accuracy was below 60% and comparable for experts and non-experts. However, experts were more consistent in their answers when the same images were repeated, and more cautious in committing to a definitive answer. Our findings demonstrate the difficulty of correctly identifying bumblebees using images from field guides. Such error rates need to be accounted for when interpreting species data, whether or not they have been collected by experts. We suggest that investigation of how experts and non-experts make observations should be incorporated into study design, and could be used to improve training in species identification.

2.2 Introduction

Accurate species identification is essential to ecological monitoring (Elphick 2008; Farnsworth et al. 2013). Species observations are used to inform and evaluate conservation actions (Sutherland, Roy & Amano 2015), such as the monitoring of population trends (Rodrigues et al. 2006; Fitzpatrick et al. 2009), the implementation and evaluation of population management plans (Duelli 1997), health assessments of ecosystems (Butchart et al. 2010), and extinction analysis (Roberts, Elphick & Reed 2010). Conversely, species *mis*identification can have serious negative impacts, such as the accidental culling of endangered species, exemplified by the endangered takahē *Porphyrio hochstetteri* (Meyer 1883) being mistaken for the destructive pukeko *Porphyrio porphyrio melanotus* (Temminck 1820) (Hunt 2015), the incorrect monitoring of harmful algal blooms (Culverhouse et al. 2003), the unobserved decline in important fish stocks (Beerkircher, Arocha & Barse 2009), and wasted resources, such as the drafting of inappropriate management plans from false species sightings (Solow et al. 2012).

While species identification in these contexts is conducted routinely by experts, such as taxonomists in museums or academic institutions (Hopkins & Freckleton 2002), there is also a long-standing tradition of members of the public supporting scientific research by contributing identification data (Stepenuck & Green 2015). Previously known as amateur naturalists, and more recently as citizen scientists, 70,000 of these lay recorders submit species observations on an annual basis in the UK alone (Pocock et al. 2015). These observers are recognised as a valuable asset in the monitoring of global environmental change (Johnson et al. 2014; Sauermann & Franzoni 2015). However, little is known about the accuracy of species identifications by non-experts, or how this compares to that of experts (Shea et al. 2011). Although some doubts have been raised over the ability of volunteers to conduct ‘real research’ (Cohn 2008), the assumption that recorded species have been correctly identified is rarely questioned (Elphick 2008). Yet the failure to account for possible species misidentification could affect assessments of population status and distribution and result in erroneous conservation decisions (Runge, Hines & Nichols 2007; Elphick 2008; Shea et al. 2011).

Few studies have investigated species identification accuracy. In a study of the classification of dinoflagellates, identification accuracy among expert observers was 72% (Culverhouse et al. 2003). Thus, more than one in four identifications was, in fact, a *mis*identification. Accuracy also varied dramatically, from 38% to 95%, depending upon the species being identified. However, accuracy was higher and more consistent in expert observers with field expertise than those with expertise gleaned from books (Culverhouse et al. 2003). In addition, individual consistency of experts with field expertise averaged 97% accuracy, but for those whose expertise came from books averaged only 75% accuracy. This indicates that observers with field experience were highly consistent in their decision-making (but both for correct and incorrect identifications), whereas the decisions of trained observers without such experience were more variable (Culverhouse et al. 2003).

A more recent study focused on the identification of individual mountain bongo antelopes *Tragelaphus eurycerus isaaci* (Thomas 1902) using a matching task (Gibbon, Bindemann & Roberts 2015). In this task, expert and non-expert observers were shown pairs of pictures of mountain bongos and had to decide whether these depicted the same or different individuals. Under these conditions, experts performed better than non-experts. However, accuracy was far from perfect in both groups of observers, with identification errors in at least one in five trials (Gibbon, Bindemann & Roberts 2015).

These results suggest that observers can be prone to identification errors in species monitoring. However, whereas one study compared different types of experts during the identification of different species (Culverhouse et al. 2003), the other compared experts and non-experts during the identification of individuals from the same species (Gibbon, Bindemann & Roberts 2015). Consequently, it is still unresolved how experts and non-experts compare directly in species identification.

In this study, we compare the identification accuracy of experts and non-experts with a matching task, in which observers have to decide whether pairs of images depict the same species. A key advantage of this task is that it allows for a direct comparison of observers with expertise in species identification with those without prior knowledge. This approach is used in other research areas, such as the study of forensic human face identification (see Johnston & Bindemann 2013) as an optimized scenario to establish best-possible performance (Burton, White & McNeill 2010) but little used in conservation research (although see Gibbon, Bindemann & Roberts (2015).

To investigate the accuracy of species identification in experts and non-experts, bees were chosen as model organisms as non-experts were likely to have experienced them, but not be overfamiliar with the different species. Using cryptic or rarely observed species groups may have deterred participants, but the range of bee species provides enough variety for testing experts too. Bumblebees (*Bombus* sp.) are generally recognisable and attractive to members of the public (Edwards & Jenner 2005), and are of great importance to human survival and the economy (Rains, Tomberlin & Kulasiri 2008; Potts et al. 2010; Klatt et al. 2014; Scriven et al. 2015). Despite this importance, bee populations are in global decline from human activities (Ghazoul 2005; Gallai et al. 2009; Williams & Osborne 2009; Potts et al. 2010; Cresswell et al. 2012). Consequently, bumblebees provide a relevant and timely model for studying the accuracy of species identification in expert and non-expert observers.

Experts and non-experts in bumblebee identification were asked to decide whether 20 pairs of bumblebee images depicted the same or two different species. To increase the relevance of this task to the monitoring of bumblebees by members of the public, the images used in this matching task were coloured illustrations of bumblebees taken from two easily accessible field guides. We sought to explore identification in detail by assessing the overall accuracy of observers in both groups, but also by exploring individual differences and the consistency of identification decisions. For this purpose, participants were asked to classify the same stimuli repeatedly, over three successive blocks.

2.3 Methods

2.3.1 Participants

This research was approved by the Ethics Committee of the School of Psychology at the University of Kent (UKC) and conducted in accordance with the ethical guidelines of the British Psychological Association. Informed consent was obtained from all participants before taking part in the survey. A total of 47 people participated in the survey, comprising expert and non-expert observers. Seven experts (3 female, 4 male, mean age = 40 years, range 25-64) were recruited via a national non-governmental organisation (NGO) specialising in the conservation of bumblebees. Forty non-experts were recruited via the School of Anthropology and Conservation at UKC (30 female, 10 male, mean age = 35 years, range = 18-65). Half of these participants (*n* = 20; 15 female, mean age = 33 years, range = 18-64) had a general background in nature conservation and were classified as non-expert conservationists (NEC). The remaining participants (*n* = 20; 15 female, mean age = 37 years, range = 18-65) had little or no experience with nature conservation and were therefore classified as non-expert non-conservationists (NENC). All 47 participants reported good vision or corrected-to-normal.

The seven expert participants reported a total of 39 years experience (1 – 15 years) in the identification of bumblebee species, whereas only seven of the non-experts reported any experience in the identification of bumblebees, ranging from 1 to 8 years. To define this experience further, all participants were asked to evaluate their identification experience on a five-point scale. Self-evaluated bumblebee identification abilities of experts and non-experts did not overlap. Non-experts reported ‘no experience’ (*n* = 33), ‘little experience’ (*n* =6) and ‘some experience’, while experts described themselves as ‘experienced’ (*n* =3, one female) and ‘competent’ (*n* = 4, two female).

2.3.2 Stimuli

The stimuli consisted of 20 pairs of images of bumblebees, comprising 10 match pairs (same species shown), and 10 mismatch pairs (different species shown), using images from two different field guides (Prŷs-Jones & Corbet, 1991; BBC Wildlife Pocket Guides, 2006). For match pairs, illustrations were from different artists (Appendix 2.3). Images in each pair consisted of colour illustrations of dorsal views of entire bumblebees, presented side-by-side on a white background. The paired images always displayed the same caste, e.g. both males, both queens etc. Stimuli were designed to be viewed on a computer monitor, and measured approximately 24 x 15 cm onscreen. No zoom function was included in the survey. Species names were taken from a checklist of extant, native bumblebees recorded in Britain and Northern Ireland (genus *Bombus* Latreille), downloaded from the Natural History Museum (London) website (www.nhm.ac.uk). For each of the species on this list, the BirdGuide application (‘app’) “Bumblebees of Britain and Ireland” (www.birdguides.com) was used to identify phenotypes associated with each species. For species that exhibited a different phenotype according to caste, an individual entry was listed for every caste that differed in appearance from other castes for each species in that guide. Although listed in the guide, *B. pomorum* (Panzer, 1805) and *B. cullumanus* (Kirby, 1802) are believed extinct, and so were removed. The randomised list also included two species in the *lucorum* complex, *B. magnus* (Vogt, 1911) and *B. cryptarum* (Fabricius, 1775), but as research shows that these are visually inseparable (Scriven et al. 2015), these were also removed. The final list comprised 45 entries representing different UK species and castes where applicable. Twenty entries were randomly sampled from the list for use in the tests. For the survey, the list of the 20 selected entries was randomised again, and the first

10 entries were the species and caste used to create match pairs. The remaining 10 entries on the list formed the first half of a mismatch pair, with the second half of the pair being selected from the other 19 species named on the list. The second species was chosen as so to create a mix of visually similar and dissimilar species (Appendix 2.3). This set of 10 match and 10 mismatch images was used to create an online survey.

2.3.3 Procedure

In the experiment, participants' bumblebee knowledge was initially recorded using two simple tasks to assess their expertise. First, participants were asked to write down all bumblebee species found in the UK (Appendix 2.1). Participants were then asked to select UK bumblebee species from a list of 40 bumblebees (20 UK and 20 non-UK species) (Appendix 2.2). On completion of the initial assessments, participants were given the matching task. In this task, participants were asked to classify each pair of bumblebees as the same species, two different species, or provide a don't know response using three different buttons on a standard computer keyboard. No time limit was applied to this task to encourage best-possible performance. Participants completed three blocks of this task. Each of these comprised the 10 match and 10 mismatch pairs, and the order of presentation was randomised for the two repeats. In the experiment, each stimulus was therefore shown three times.

Analyses

Participant knowledge of bumblebee species was calculated from questionnaire answers. Percentage accuracy was calculated for each participant, and responses were analysed within the three *a priori* expert groups, Mean accuracy for each of the three groups was presented for every trial, match trials, and mismatch trials. Correct, incorrect and don't know responses were also presented within the expert groups. Accuracy was then calculated for each of the three separate trials to measure whether participants were consistent in their answers for the same image. Using SPSS (IBM SPSS Statistics for Macintosh, Version 23.0) one way ANOVAs compared accuracy and consistency between expert groups, and a mixed-factor ANOVA compared expert groups and trial type (match, mismatch). Arcsine square-root transformed data were used for ANOVAs, to normalise the proportional data. A Pearson product-moment correlation was used to investigate relationship between experience and accuracy.

2.4 Results

2.4.1 Participant expertise

Experts had substantial knowledge of bumblebees, whereas the non-experts knowledge was minimal. On average, experts named 20.7 bumblebee species (SD = 4.8; min = 15; max = 25), non-expert conservationists (NECs) named on average only 0.4 species (SD = 0.8; min = 0; max = 3), and non-expert non-conservationists (NENCs) only 0.2 species (SD = 0.4; min = 0; max = 1). Similarly, experts correctly chose an average of 19.7/20 UK species from a list of 40 *Bombus* species (Appendix 2.2) (SD = 0.5; min = 19; max = 20), whereas NECs could only select an average of 1.6 species (SD = 2.1; min = 0; max = 7) and NENCs only 0.1 species (SD = 0.3; min = 0; max = 1).

2.4.2 *Bee matching accuracy*

Overall accuracy in the matching task was low and similar across groups of expertise (Fig. 2.1), with the mean percentage of correct responses ranging from 54% to 57%. Correspondingly, incorrect responses were high and recorded on between 33% (for experts) and 42% (NEC) of trials across groups. Finally, experts made don't know responses on 11% of trials, while this contributed to less than 5% of responses in both groups of non-experts.

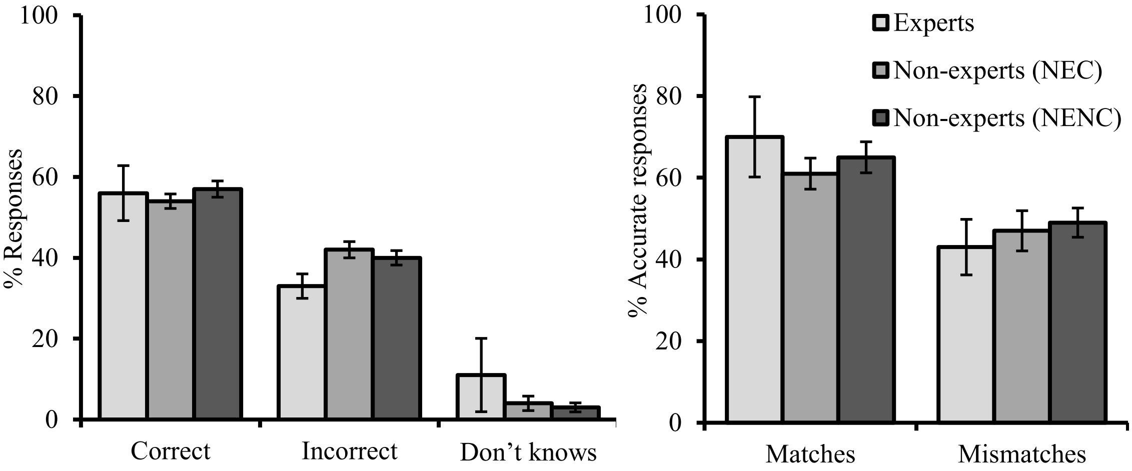


Figure 2.1. Percentage (± 1 s.e.) for correct, incorrect and don't know responses (left graph), and accuracy (± 1 s.e.) for match and mismatch pairs (right graph) as a function of expertise. Overall accuracy is low and comparable (54% to 57%) between expert groups.

There was no difference between the three participant groups (E, NEC, and NENC) in terms of correct responses ($F(2,44) = 0.45$, $p = 0.638$), incorrect responses ($F(2,44) = 2.89$, $p = 0.066$), and don't know responses ($F(2,44) = 0.35$, $p = 0.704$). Thus, experts and non-experts overall accuracy did not differ on this task. Match performance was similar across expertise groups, at between 61% and 70%

accuracy, while mismatch performance was generally lower, at between 43% and 49% accuracy, but similar across the groups (Fig. 2.1). In line with these observations, a 3 (group: E, NEC, NENC) x 2 (trial type: match, mismatch) mixed-factor ANOVA found an effect of trial type ($F(1,44) = 13.50$, $p = 0.001$), but not of expertise, ($F(2,44) = 0.61$, $p = 0.545$), and no interaction between factors ($F(2,44) = 0.59$, $p = 0.557$).

2.4.3 Experience and matching accuracy

Percentage accuracy of correct, incorrect and don't know responses were correlated with the years of experience that all participants reported in the identification of UK bumblebees. A Pearson product-moment correlation conducted on responses showed that correct and incorrect responses declined with experience ($r = -0.27$, $n = 47$, $p = 0.072$ and $r = -0.30$, $n = 47$, $p = 0.038$, respectively), but don't know responses increased with experience ($r = 0.54$, $n = 47$, $p < 0.001$). This suggests that the more experienced observers were less likely to commit to a correct or incorrect identification decision. This inference is drawn tentatively, considering the limited sample of experts and possible extreme scores in the data (see *Participants* section).

2.4.4 Experience and accuracy for individual items

Accuracy was also calculated for all individual stimulus pairs and the groups of observers. For this by-item analysis, accuracy was combined across the three presentations of each stimulus pair (Fig. 2.2). One factor ANOVAs for each match and mismatch stimulus show that effects of expertise were present for only three of the 20 images. Post-hoc Tukey tests reveal that experts outperformed non-experts with Match 2 ($F(2,44) = 5.92$, $p = 0.005$; E v NEC and E v NENC $p = 0.007$) and

Match 6 ($F(2,44) = 3.76$, E v NEC $p = 0.024$). Conversely, non-experts outperformed experts with Mismatch 7 ($F(2,44) = 7.00$, NEC v E $p = 0.005$ and NENC v E $p = 0.002$).

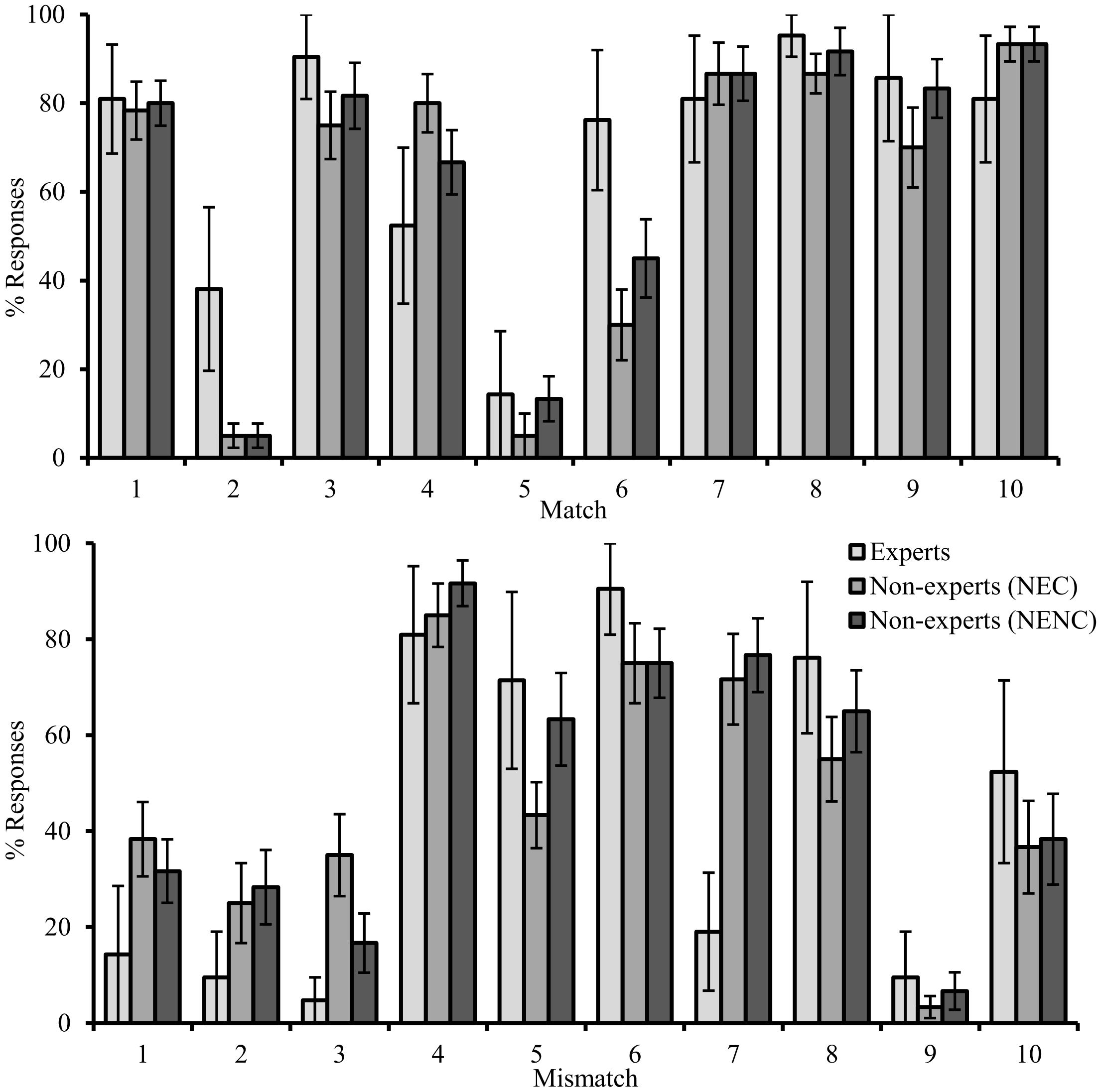


Figure 2.2. Mean accuracy (± 1 s.e.) across groups for each match (top) and mismatch (bottom) image. Effects of expertise were present for only three of these images (match 2, match 6 and mismatch 7).

This pattern suggests that the reliance on purely visual information by non-experts generally leads to comparable and occasionally even better accuracy than experts. This indicates that the additional subject-specific experience of experts does not consistently improve, and might even hinder, the matching of some bumblebee species. In some cases, however, expert knowledge also adds a performance advantage that must transcend the available visual information.

2.4.5 Consistency

We also sought to determine whether experts might be more consistent than non-experts in their identification of bumblebees, by assessing performance across the three repeated trials. Consistent decisions were defined as instances in which observers made the same responses to bumblebee pairs in all three trials. Two consistency measures were obtained. The first of these reflects overall consistency regardless of accuracy, and was calculated by collapsing consistent correct (42% of all decisions), incorrect (27%), and don't know (2%) responses for the different expertise groups (Fig. 2.3).

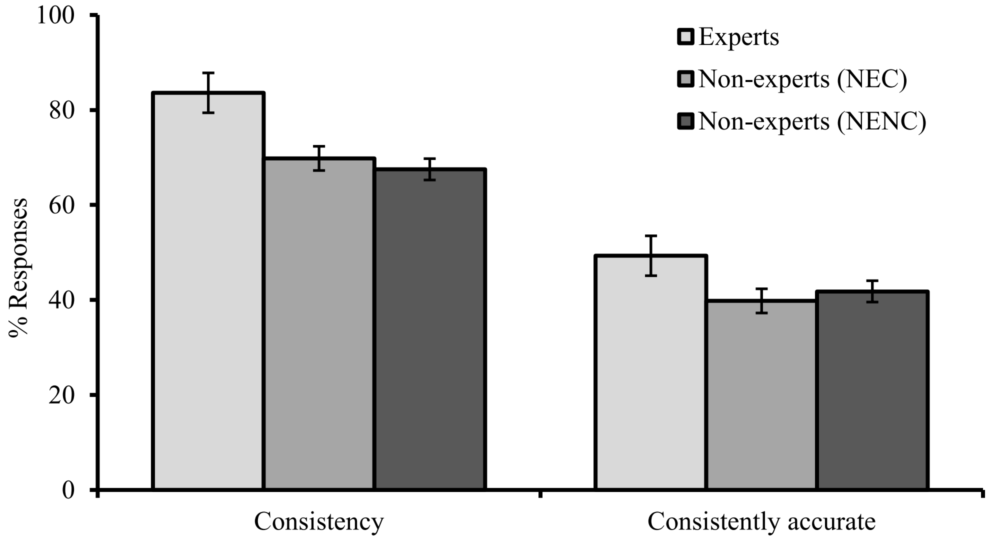


Figure 2.3. Percentage (± 1 s.e.) consistency in responses across presentations of stimuli. Experts were more consistent than non-experts in both overall and accurate answers.

A one-factor ANOVA showed that consistency varied between participant groups (*F*(2,44) = 5.42 *p* = 0.008). Tukey post-hoc test showed that experts were more consistent than the NEC (*p* = 0.020) and the NENC groups (*p* = 0.006). The consistency of NECs and NENCs did not differ (*p* = 0.830). A second consistency measure was calculated, which reflects the consistency of accurate responses only. This revealed a similar pattern, with experts outperforming the two non-expert groups. A one-factor ANOVA showed that these differences between groups were not reliable (*F*(2,44) = 0.55, *p* = 0.583). However, a Pearson product-moment correlation between consistent and consistently-accurate responses (Fig. 2.4) was found (*r* = 0.722, *n* = 47, *p* < 0.001). Taken together, these data indicate that experts are generally more consistent than non-experts in their responses, but not in their accurate responses. However, the individuals (experts or non-experts) whose responses are more consistent are also more likely to be consistently accurate.

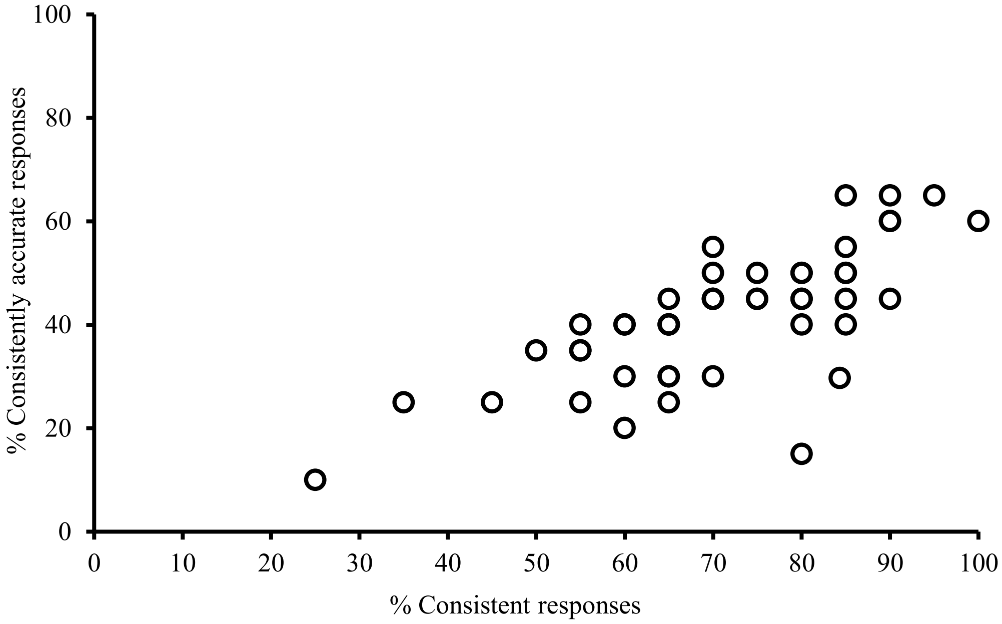


Figure 2.4. Correlation of consistent and consistently accurate responses. Individuals with consistent responses are more likely to be consistently accurate.

2.5 Discussion

This study examined experts' and non-experts' accuracy in the identification of bumblebees. The naming tasks revealed clear effects of expertise, with experts naming an average of 20.7 UK bumblebee species and selecting 19.7/20 from a list of bumblebee species, while NECs and NENCs could name on average less than one species and selected less than two. A different picture emerged when the *actual* visual identification accuracy of these observers was assessed with the matching task. In this task, experts' overall accuracy was low, at 56%, and indistinguishable from NECs and NENCs. This finding was confirmed when performance was broken down into match and mismatch trials, for which expert and non-expert performance also did not differ. Participants' self-reported years of experience in bumblebee identification was also correlated with responses on the matching task. This analysis shows that both correct and incorrect responses decline with experience, but don't know responses increase. Thus, observers appear to become more cautious with experience and less willing to commit to any identification decisions. These inferences are drawn tentatively, due to the limited availability of bumblebee experts for this study (n = 7). Crucially, however, these findings suggest once again that expertise does not improve the visual identification of bumblebees in the present task.

Overall, these findings converge with studies that have shown that visual species identification can be surprisingly error-prone. In contrast to previous studies, which either examined different types of experts during the identification of different species (Culverhouse et al. 2003), or experts and non-experts during the identification of individuals from the same species (Gibbon, Bindemann & Roberts 2015), the current experiment compared experts and non-experts during species identification. The error rates that are observed across these studies raise important questions concerning the accuracy of species identification using field guides. Such identifications are used for supporting a wide range of actions, such as the monitoring of endangered species (Barlow et al. 2015; Dennhardt et al. 2015; Lee et al. 2015) and the drafting of appropriate management plans (Guisan et al. 2013; Tulloch et al. 2013; Lukyanenko, Parsons & Wiersma 2011). An understanding of error rates needs to be factored into such important conservation activities.

We draw these conclusions with some caveats. It is conceivable, for example, that accuracy among experts and non-experts will vary depending on how images are presented or which guidebooks are at hand. A number of identification guides exist for UK bumblebees, providing a variety of pictures of bumblebees, e.g. line or colour drawings, photographs and stylised diagrams. The extent to which illustrations from different guides accurately capture the key visual features of different bumblebee species and also match each other remains open to exploration (Fitzsimmons 2013), but is bound to affect bumblebee identification tasks. Variation in the specimens used by illustrators may also be due to phenotypic variation or even mislabelling in the collections used. Moreover, a difficult question for illustrators is which individual of a species should be drawn in order for a guidebook to represent a ‘typical’ specimen. There is the option to use the holotype (the single specimen on which the species is described), but this may not be readily available or representative of the current UK population. There is the added complication of physical differences caused by age, such as the loss of hair and fading of colour due to the sun (Edwards & Jenner 2005).

There is a long history of members of the public contributing to species monitoring programmes (Bonney et al. 2009; Silvertown 2009; Sutherland, Roy & Amano 2015). An analysis of accuracy for individual items indicated that expert and non-expert accuracy was similar for most bumblebee species here. There were also instances in which experts outperformed non-experts or the reverse pattern was found. This indicates that for some species comparisons, the reliance on purely visual information (as available to both experts and non-experts) produces best accuracy. In other cases, the additional subject-specific expert knowledge can occasionally interfere with the visual identification process. However, expert knowledge can also transcend the available visual information in some cases and provide a benefit in performance. This mixture of results is an intriguing outcome that is perhaps counter to intuition, because it suggests that the identification accuracy of bumblebees might be optimized best by using expert and non-expert decisions in a complementary fashion.

Experts were more consistent in their decisions when the tests were repeated. This effect was only reliable when correct, incorrect and don’t know decisions were combined. Overall, however, the more consistent observers were also more consistently accurate. Thus, experts’ decision criteria appear to be more stable and this might confer an advantage when identification of the same species is assessed repeatedly. Further, systematic investigations of these different effects (visual vs. expertise-driven identifications, consistency) might inform training that is designed to enhance the accuracy of observers. This could help to reduce error rates and improve species monitoring in the field.

Additional identification cues might also be available that could specifically enhance expert performance in practical settings, such as context, behaviour, flight period, or even the presentation of a live or dead specimen. Experts will also have access to additional resources to support identification, such as taxonomic revisions with identification keys and diagrams or natural history collections. We included a short questionnaire in our study to assess field guide usage, which showed that all seven experts reported a combination of up to five field guides, and three also utilised smartphone apps. However, our data suggests that this experience did not enhance performance in the current experiment. More generally, it remains unresolved whether sufficient numbers of experts can be found for research to provide the volumes of data required to understand such factors (Kelling et al. 2015). Results may also differ for other taxa, but the growth of citizen science and the increase in use of these volunteer data means that species observations, such as those used to inform conservation practitioners, are likely to be heavily reliant on images, either as submissions by non-experts or validation by experts.

In conclusion, this study shows that experts and non-experts both make many errors when using standard field guide illustrations to identify species. This raises important questions surrounding the accuracy of species observations in ecological datasets, and suggests that consideration should be given to possible inaccuracies when such information is used to inform decision makers.

2.6 References

Barlow, K.E., Briggs, P.A., Haysom, K.A., Hutson, A.M., Lechiara, N.L., Racey, P.A., Walsh, A.L. and Langton, S.D. (2015). Citizen science reveals trends in bat populations: The National Bat Monitoring Programme in Great Britain. *Biological Conservation*, **182**, 14–26.

BBC Wildlife Pocket Guides Number 11 (2006). *Bumblebees other bees and wasps*. BBC Wildlife Magazine, Bristol.

Beerkircher, L., Arocha, F. and Barse, A. (2009). Effects of species misidentification on population assessment of overfished white marlin *Tetrapturus albidus* and roundscale spearfish *T. georgii*. *Endangered Species Research*, **9**, 81-90.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. and Shirk, J. (2009). Citizen Science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, **59,** 977-984.

Burton, A.M., White, D. and McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, **42**, 286-291.

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E. (2010). Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Cohn, J.P. (2008). Citizen Science: can volunteers do real research? *BioScience*, **58**, 192-197.

Cresswell, J.E., Page, C.J., Uyguna, M.B., Holmbergh,M., Li, Y., Wheeler, J.G, Laycock, I., Pook, C.J., Hempel de Ibarrac, N., Smirnoff, N. and Tyler, C.R. (2012). Differential sensitivity of honey bees and bumble bees to a dietary insecticide (imidacloprid). *Zoology*, **115**, 365–371.

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, **247**, 17–25.

Dennhardt, A.J., Duerr, A. E., Brandes, D. and Katzner, T. E. (2015). Integrating citizen-science data with movement models to estimate the size of a migratory golden eagle population. *Biological Conservation*, **184**, 68–78.

Duelli, P. (1997). Biodiversity evaluation in agricultural landscapes: an approach at two different scales. *Agriculture, Ecosystems & Environment*, **62**, 81–91.

Edwards, M. and Jenner, M. (2005). *Field guide to the bumblebees of Great Britain and Ireland*. Ocelli Ltd, Eastbourne.

Elphick, C.S. (2008). How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, **45**, 1313–1320.

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J. Stevenson, R.D., Courtney, G.W., Van Dyk, J.K. and Ellison A.M. (2013). Next-generation field guides. *BioScience*, **63**, 891–899.

Fitzpatrick, M.C., Preisser, E. L., Ellison, A. M. and Elkinton, J. S (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, **19**, 1673–1679.

Fitzsimmons, J.M. (2013). How consistent are trait data between sources? A quantitative assessment. *Oikos*, **122**, 1350-1356.

Gallai, N., Salles, J-M., Settele, J. and Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, **68**, 810–821.

Ghazoul, J. (2005). Buzziness as usual? Questioning the global pollination crisis. *Trends in Ecology & Evolution*, **20**, 367–373.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ*, **3**, e1303.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R. Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield,S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. and Buckley, Y.M. (2013). Predicting species distributions for conservation decisions. *Ecology letters*, **16**, 1424–1435.

Hopkins, G.W. and Freckleton, R.P. (2002). Declines in the numbers of amateur and professional taxonomists : implications for conservation. *Animal Conservation*, **5**, 245–249.

Hunt, E. (2015). New Zealand hunters apologise over accidental shooting of takahē. *The Guardian* [Online], (Environment) 21 August. Available from: http://www.theguardian.com/environment/2015/aug/21/new-zealand-conservationists-apologise-over-accidental-shooting-of-endangered-takahe [Accessed 21st August 2015].

Johnson, M.F., Hannah, C., Acton, L., Popovici, R., Karanth, K.K. and Weinthal, E. (2014). Network environmentalism: Citizen scientists as agents for environmental advocacy. *Global Environmental Change*, **29**, 235–245.

Johnston, R.A. and Bindemann, M. (2013). Introduction to forensic face matching. *Applied Cognitive Psychology*, **27**, 697–699.

Kelling, S., Fink, D., La Sorte, F. A., Johnston, A., Bruns, N. E. and Hochachka, W. M. (2015). Taking a ‘Big Data’ approach to data quality in a citizen science project. *Ambio*, **44**, 601–611.

Klatt, B.K., Holzschuh, A., Westphal, C., Clough, Y., Smit, I., Pawelzik, E. and Tscharntke, T. (2014). Bee pollination improves crop quality, shelf life and commercial value. *Proceedings of the Royal Society B*, **281**, 20132440.

Lee, T.E., Black, S. A., Fellous, A., Yamaguchi, N., Angelici, F. M., Al Hikmani, H., Reed, J. M., Elphick, C. S. and Roberts, D. L (2015). Assessing uncertainty in sighting records: an example of the Barbary lion. *PeerJ*, **3**, e1224.

Lukyanenko, R., Parsons, J. and Wiersma, Y. (2011) Citizen Science 2.0: data management principles to harness the power of the crowd. Hemant, J., Sinha, A.P. and Vitharana P., eds. *Service-Oriented Perspectives in Design Science Research*, Springer, Berlin. 465-473.

Pocock, M.J.O., Roy, H. E., Preston, C. D. and Roy, D. B. (2015). The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnean Society*, **115**, 475–493.

Potts, S.G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution*, **25**, 345–353.

Prŷs-Jones, O.E. and Corbet, S.A. (1991) *Bumblebees*. Richmond Publishing Company Limited, London.

Rains, G.C., Tomberlin, J.K. and Kulasiri, D., (2008). Using insect sniffing devices for detection. *Trends in biotechnology*, **26**, 288–294.

Roberts, D.L., Elphick, C.S. and Reed, J.M. (2010). Identifying anomalous reports of putatively extinct species and why it matters. *Conservation biology*, **24**, 189–196.

Rodrigues, A.S.L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M. and Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in Ecology and Evolution*, **21**, 71–76.

Runge, J., Hines, J. and Nichols, J. (2007). Estimating species-specific survival and movement when species identification is uncertain. *Ecology*, **88**, 282-288.

Sauermann, H. and Franzoni, C. (2015). Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*. **112**, 679-684.

Scriven, J.J., Woodall, L. C., Tinsley, M. C., Knight, M. E., Williams, P. H., Carolan, J. C., Brown, M. J. F., Goulson, D. (2015). Revealing the hidden niches of cryptic bumblebees in Great Britain: implications for conservation. *Biological Conservation*, **182**, 126–133.

Shea, C.P., Peterson, J. T., Wisniewski, J. M. and Johnson, N. A. (2011). Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, **24**, 467–471.

Solow, A., Smith, W., Burgman, M., Rout, T., Wintle, B. and Roberts, D. (2011). Uncertain sightings and the extinction of the ivory-billed woodpecker. *Conservation Biology*, **26**, 180–184.

Stepenuck, K. and Green, L. (2015). Individual-and community-level impacts of volunteer environmental monitoring: a synthesis of peer-reviewed literature. *Ecology and Society*, **20**, 19.

Sutherland, W.J., Roy, D.B. and Amano, T. (2015). An agenda for the future of biological recording for ecological monitoring and citizen science. *Biological Journal of the Linnean Society*, **115**, 779–784.

Tulloch, A.I.T., Possingham, H. P., Joseph, L. N., Szabo, J. and Martin, T. G. (2013). Realising the full potential of citizen science monitoring programs. *Biological Conservation*, **165**, 128-138.

Williams, P.H. and Osborne, J.L. (2009). Bumblebee vulnerability and conservation world-wide. *Apidologie*, **40**, 367–387.

Appendix 2.1

Many thanks for taking part in this experiment, conducted as part of a doctoral research project at the University of Kent.

If you would like to be contacted regarding the findings of this study, please express your interest by emailing gea7@kent.ac.uk.

Question 1.

Can you name any species of bumblebee found in the UK?

(Either common or Latin name is acceptable)

Appendix 2.2

Question 2.

Please highlight (e.g. circle, strikethrough, change colour) any of the below bumblebee (*Bombus*) species that you believe occur in the UK:

argillaceus
armeniacus
atripes
barbutellus
bohemicus
campestris
cryptarum
cullumanus
deuteronymus
fragrans
hedini
hortorum
humilis
hyperboreus
hypnorum
hypocrita
jonellus
ladakhensis
lapidarius
lucorum
margreiteri
monticola
morawitzianus
muscorum
opulentus
pascuorum
perezi
persicus
ruderarius
ruderatus
rupestris
schrencki
soroeensis
subterraneus
sylvarum
sylvestris
terrestris
tichenkoi
ussurensis
wurflenii

Appendix 2.3

A table showing the *Bombus* species and caste, with the relevant illustrator, for each pairwise image used in this task.

Image Type	**Species and caste**	**Species and caste**	**Artist**	
			Left image	**Right image**
Match	*B. distinguendis* (Queen)		C. Shields	A.J. Hopkins
Match	*B. hortorum* (Queen)		C. Shields	A.J. Hopkins
Match	*B. humilis* (Queen)		A.J. Hopkins	C. Shields
Match	*B. jonellus* (Queen)		C. Shields	A.J. Hopkins
Match	*B. pratorum* (Male)		C. Shields	A.J. Hopkins
Match	*B. ruderarius* (Male)		C. Shields	A.J. Hopkins
Match	*B. ruderarius* (Queen)		A.J. Hopkins	C. Shields
Match	*B. soroensis* (Queen)		A.J. Hopkins	C. Shields
Match	*B. sylvarum* (Queen)		C. Shields	A.J. Hopkins
Match	*B. terrestris* (Queen)		A.J. Hopkins	C. Shields
Mismatch	*B. hortorum* (Queen)	*B. ruderatus* (Queen)	C. Shields	C. Shields
Mismatch	*B. lapidarius* (Male)	*B. ruderarius* (Male)	A.J. Hopkins	A.J. Hopkins
Mismatch	*B. lapidarius* (Queen)	*B. ruderarius* (Queen)	C. Shields	C. Shields
Mismatch	*B. lucorum* (Queen)	*B. terrestris* (Queen)	A.J. Hopkins	C. Shields
Mismatch	*B. lucorum* (Queen)	*B. soroensis* (Queen)	C. Shields	C. Shields
Mismatch	*B. monticola* (Queen)	*B. lapidarius* (Queen)	A.J. Hopkins	A.J. Hopkins
Mismatch	*B. muscorum* (Male)	*B. humilis* (Male)	C. Shields	C. Shields
Mismatch	*B. pascuorum* (Male)	*B. muscorum* (Male)	C. Shields	C. Shields
Mismatch	*B. pascuorum* (Queen)	*B. sylvarum* (Queen)	A.J. Hopkins	C. Shields
Mismatch	*B. subterraneus* (Queen)	*B. monticola* (Queen)	A.J. Hopkins	C. Shields

Sloth or pain au chocolat? ©teenybiscuit

Chapter 3.

3 Inconsistency in species identification under optimized conditions

Revised following peer review from *Scientific Reports* SREP-16-42881A
Currently in review

3.1 Abstract

Species identification is the procedure by which an unknown organism is assigned to a species. For those without prior knowledge, deciding which species has been observed often involves the use of visual aids, such as keys, field guides or voucher specimens. This may appear to be a simple decision-making process and is rarely questioned. In this study, non-experts performed a series of matching tasks that involved images of 20 UK bumblebee queens being paired with each other. Accuracy varied greatly, from 11% to 100% for species mismatches. We found inconsistency in accuracy across the different species combinations, as well as a mixture in agreement with similar species as recommended by an identification guide. We found that neither the identification ability of novice observers nor inter-species variability to be uniformly dissimilar. These findings may help with the design of training aids for novices, such as citizen science projects, and suggest that the choices of 'ideal' focal species in surveys may minimise the chance of misidentifications.

3.2 Introduction

The recording of the natural world is a practice that can be traced back over centuries (Cole 1944; Allen 1976; Lawrence 2009; Miller-Rushing 2012). Much of this activity has been conducted by amateur naturalists, who have made a significant contribution to our understanding of species occurrence (Lawrence 2009; Gardiner et al. 2012; Tulloch et al. 2013; Merenlender and Crall 2016). The growing popularity of citizen science, in addition to volunteers already monitoring biodiversity, has also resulted in individuals developing skills in species identification (Kremen, Ullman and Thorp 2011; Farnsworth et al. 2013). These activities include both ‘top-down’ recording, where projects are designed for volunteers to collect specific data, and ‘bottom-up’ recording, when observations serve the purposes of the recorder. Either way, both amateur naturalists and citizen scientists are often self-taught and autonomous (Curtis 2015; Robinson, Inger and Gaston 2016), which sometimes raises questions about the accuracy of the records these observers generate (Fore, Paulsen and O’Laughlin 2001; Cohn 2008; Crall et al. 2011; Gollan et al. 2012).

Whether identifying species independently or as part of a specific project, there are a number of methods that may be used in this process, such as DNA barcoding, dichotomous keys, voucher specimens, prior knowledge and field guides. In practice, not all of these methods will be widely known or accessible, and the practicality of printed, illustrated field guides makes them the traditional tool for identifying unknown organisms (Farnsworth et al. 2013). Using a field guide is predominantly a visual process that entails matching a sighting, a specimen or an image with one of a selection of photos or illustrations, usually accompanied by information about that taxa. Traditionally, field guides have been utilitarian, and are credited with enabling learning through a combination of visual and written information (Givens, Reeds and

Touwaide 2006; Scharf 2009). The basic structure of a guide comprises a series of illustrations of species believed to occur in a geographical area or taxon group, with an accompanying description (Stevenson, Haber and Morris 2003; Dunlap 2005; Farnsworth et al. 2013). The identification process involves the user scanning the available images, deciding whether any of those images match the unknown organism, and naming that organism accordingly. Within the species description, many guides include species believed to be similar in appearance to the target species to help focus the identification process.

Correctly identifying species is vital to ecological monitoring (Elphick 2008; Farnsworth et al. 2013). Important and costly decisions are made on species numbers and distributions (Rodrigues et al. 2006; Butchart et al. 2010; Sutherland, Roy and Amano 2015; Juffe-Bignoli et al. 2016), and the possible consequences of *mis*identification can be disastrous, such as misunderstanding the impacts of agricultural practices (Duelli 1997), commercial fishing (Beerkircher, Arocha and Barse 2009; Tillett et al. 2012), and invasive species (Fitzpatrick et al. 2009). The implications of misidentification may also apply to existing datasets or collections, which provide historical baselines for species and their distribution (Ellwood et al. 2016). A recent study found, for example, that over 50% of tropical specimens in herbaria are likely to be incorrectly named (Goodwin et al. 2015). Museum specimens are often used for reference, both in-collection and via remote access, which could result in such misinformation being spread widely (Goodwin et al. 2015).

Whilst such studies raise concern about the accurate naming of plants, the number of described insects is threefold (Goodwin et al. 2015) and consequently, invertebrate misidentification could have an even greater impact. In terms of conservation decision-making, without reliable evidence of species distributions the assessment of certain species conservation status is difficult, and subsequently determining where to focus conservation efforts could be misguided (Zapponi et al. 2016). In this context, one taxon of great conservation concern are bees, especially given their important role in the ecosystem service of pollination (Biesmeijer et al., 2006). Even though many bee species, including common ones, look very similar to each other, they have different life histories and perform different roles in the environment (Paxton et al. 2015). For example, bumblebees (*Bombus spp.*) are particularly important for the pollination of wild plants. For some plants, pollination is performed by just one species of bumblebee, and a decline in bumblebees could be linked to the loss of specific habitats (Goulson et al. 2005). However, insufficient monitoring means that patterns of decline are difficult to determine (Casey et al. 2015). Moreover, this information could be impeded by species misidentifications that are unaccounted for.

Bumblebees are generally recognisable and attractive to members of the public (Edwards and Jenner 2005), and consequently not an unfamiliar group of insects. In the UK alone there are over 20 different species, with distinctive colour patterns being their most obvious feature (Williams, 2007). The inter-species similarity of these taxa makes them good model species by which to study character and regional variation in colour patterns (Williams, 2007). However, inter- and intra-species morphology can be highly variable (Cameron et al., 2006). For example, castes of some species exhibit the same colours and patterns, while in other species, each of the three castes (queen, male and worker) differ morphologically (see Edwards and Jenner 2005). With inter-species variation, some species, such as the '*lucorum*' complex, are so similar in their morphology that they can only be distinguished genetically (Scriven et al. 2015). Not only does this raise concerns over the true numbers and distribution of the species in this cryptic group, but also whether existing ecological information applies to just one or several different species (Scriven et al. 2015).

Using a pairwise matching task under highly optimised conditions, the current study examined the ability of non-expert participants to identify whether two images of bumblebees were the same or different species. In this task, novice observers were shown two simultaneously presented images of bumblebees taken from an established identification guide. As well as inviting participants that were unfamiliar with bees, stimuli comprised images of bumblebees on a white background, reducing influence of prior knowledge and context, with the aim of decisions being based on visual information alone. This method provides an initial step towards a direct, baseline measure of the accuracy of species identification. In the absence of prior training and additional information, this approach can provide direct and objective insight into the visual confusability of species by human observers (Williams, 2007). Other studies have found heterogeneity in identification by individuals at an intra-species level (Gibbon, Bindemann and Roberts 2015), and between groups of individuals according to expertise at an inter-species level, for which no effect of expertise was found (Austen, Bindemann and Roberts 2016). Here we investigated differences in accuracy across species within the same genus. Specifically, we explored which species were associated with low accuracy rates, whether certain species are perceived as more similar than others, and if there are any distinguishing visual features associated with higher levels of similarity.

3.3 Method

3.3.1 Participants

This research was approved by the Ethics Committee of the School of Psychology at the University of Kent and conducted in accordance with the ethical guidelines of the British Psychological Association. A total of 104 people participated in the survey (93 female, 11 male, mean age = 24.9 years, SE = 0.8). These were recruited by email and social media via the Durrell Institute of Conservation and Ecology (DICE) and the School of Psychology at the University of Kent (UKC) and received compensation for their time in the form of a £10 voucher or course credit. Observers reported their age on interval scales, and the mid-point of each category was used to calculate observers' mean age. All participants reported good vision or corrected-to-normal vision with glasses or contact lenses.

3.3.2 Stimuli

Stimuli contained images of queens from 20 bumblebee species found in the UK. These model species were taken from a list of extant, native bumblebees recorded in Britain and Northern Ireland (genus *Bombus*), downloaded from the Natural History Museum of London's website. This list was randomised, and the first 20 species were used as stimuli. Using colour illustrations of dorsal views of entire queens taken from popular bumblebee identification guides (Prŷs-Jones and Corbet, 1991; BBC Wildlife Pocket Guides, 2006), each species was systematically paired with each of the other 19 species. Presented side-by-side on a white background, this yielded 210 image pairs. These comprised 20 match pairs, in which the same species is shown in the two images, and 190 mismatch pairs, in which two different species were shown.

3.3.3 Procedure

Using the polling software SurveyGizmo (surveygizmo.com), participants anonymously provided demographic information. To verify that participants were unfamiliar with the target species, they were asked about their experience with bumblebee identification and their use of identification guides. This was followed by a matching task, for which the image order was randomised for each participant. Images appeared consecutively and participants were asked to respond using three different buttons on a standard computer keyboard. For each image, participants were asked whether they thought that the two images in the pair were a match, with the option of "Yes", "No" or "Don't know" in order to avoid a forced choice. No time limit was applied to this task to encourage best-possible performance (Özbek and Bindemann 2011)

3.3.4 Analyses

Individual participant performance was analysed overall, and also for match (same species) pairs and mismatch (different species) pairs. Species combinations were also analysed. For match pairs (*n* = 20), accurate responses were reported as percentages for each species. For mismatch pairs (*n* = 190), results were analysed using both accurate and inaccurate scores. Accuracy for mismatch pairs means that participants correctly identified that the image showed two different species. Inaccurate scores reflect participants misidentifying the mismatch pair as the same species when they are actually different species. This 'misidentification rate' excludes don't know answers, therefore giving a true reflection of inter-species misidentification. Accuracy was also analysed per species and by colour. To analyse accuracy by species, a mean accuracy score was calculated for each of the 20 species, by averaging the accuracy for each image featuring that particular species. We also analysed this data by 'tail' colour, Tail colour refers to the colour of the hairs on the posterior part of the abdomen and is a defining visual feature of bumblebees. The pattern is usually bright with strong contrasts in colour (Williams 2008), and as such is a common descriptor, often defining how bumblebees are grouped in field guides. Difference in accuracy between tail colours was tested with t-tests and one-way ANOVAs (SPSS version 24). 'Confused' species were also compared with the findings from this matching study. Some field guides note species that are easily confused. The species noted as most confused in the species descriptions in Falk (2015), were compared those most confused in this study. Inferential statistics were performed using arcsine square-root transformed data as the data are proportions which truncate at 0 and 1, and the data is required to be asymptotic.

3.4 Results

3.4.1 Self-reported experience in bumblebee identification

Participants as novel observers was supported, as the majority had no experience with bumblebee identification (*n* = 84), and the remainder reported a little experience but not to a species level (*n* = 15), some experience with identifying common species (*n* = 4), and experience beyond the common species (*n* = 1). Some guides define certain bumblebee species as 'common' or 'rare' (e.g. Edwards & Jenner, 2005; Pilchen, 2010), but the term common was not defined for participants in their self-assessment of expertise. None reported being competent in identification, with only 9% having used guides specifically designed for bumblebee identification.

3.4.2 Accuracy

The average for correct responses across all images was 78.4% (range = 19-100%, SE = 1.4), due to 98.1% accuracy for match trials (range = 70-100%, SE = 0.5) and 76.3% for mismatch trials (range = 11-100%, SE = 1.5) (Fig. 3.1a).

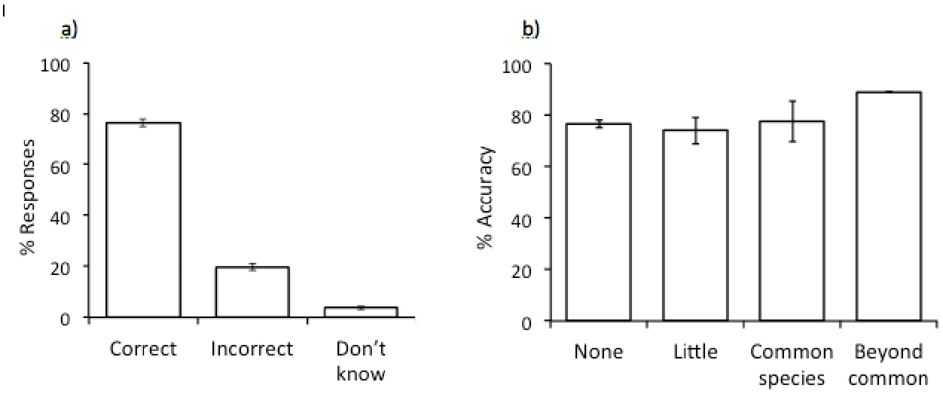


Figure 3.1. Percentage participant accuracy for mismatch pairs, showing correct, incorrect and don't know responses (a) and average accuracy by experience (b).

To analyse mismatch accuracy by species, a mean accuracy score was calculated for each of the 20 species, by averaging the accuracy for each mismatch image featuring that particular species. This data is provided in Table 3.1 and shows enormous inter- and intra- species variation. For example, mean accuracy across species ranges from 92% for mismatch images including *B. hypnorum* to 63% for those including *B. subterraneus*. In addition, *B. hypnorum* also showed the minimum intra-species variation (13%), but it did not follow that the maximum intra-species variation was found in *B. subterraneus* (67%). Maximum variation was within *B. pascuorum* and *B. muscorum*, and was larger at 93% for both species. These data suggest that accuracy in species identification is extremely variable between participants, despite the similar lack of familiarity with the target species, and accuracy is highly dependent upon the combination of species being compared.

The most accurate and least varied results were for mismatch images containing *B. hypnorum* (Table 3.1), which is of interest as this species has the unique colour combination of a ginger thorax, no stripes, and a white tail. All other white tail bumblebees are black with combinations of yellow stripes. Although the lowest *mean* accuracy was found for images containing *B. subterraneus*, the lowest accuracy for a mismatch in this species group was 25%, which is considerably greater than the lowest overall accuracy score of 4% (Table 3.1). The species most confused with *B. subterraneus* was *B. sylvarum*, which also has two yellow stripes on the thorax, but a different tail colour (see Fig. 3.2). The lowest accuracies (4%) and the largest variations (93%) were for *B. pascuorum* and *B. muscorum,* but mean accuracy was 72% in both species groups, placing them in the middle of Table 3.1.

Table 3.1. Participant accuracy scores for mismatch pairs reported by species shown in descending order of the mean. The minimum and maximum values are the lowest and highest accuracy scores for a mismatch pair that included that species. The corresponding species for those pairs are named in parentheses.

Species	Min	Max	Mean	SD
B. hypnorum	85.6% (*B. barbutellus, B. bohemicus, B. soroensis*)	98.1% (*B. lapidarius*)	91.5%	± 4.1
B. lucorum	64.4% (*B. soroensis*)	98.1% (*B. monticola*)	86.9%	±10.1
B. rupestris	47.1% (*B. lapidarius*)	96.2% (*B. soroensis*)	86.6%	±12.2
B. bohemicus	69.2% (*B. barbutellus*)	97.1% (*B. muscorum B. ruderarius*)	86.3%	± 9.1
B. barbutellus	55.8% (*B. ruderatus*)	97.1% (*B. pascuorum*)	86.0%	±12.0
B. ruderarius	17.3% (*B. lapidarius*)	97.1% (*B. barbutellus, B. bohemicus*)	83.2%	±18.1
B. terrestris	62.5% (*B. pratorum*)	96.2% (*B. rupestris*)	82.9%	± 9.9
B. lapidarius	17.3% (*B. ruderarius*)	98.1% (*B. hypnorum*)	81.6%	±19.7
B. monticola	45.2% (*B. distinguendus*)	98.1% (*B. bohemicus*)	77.7%	±14.9
B. ruderatus	32.7% (*B. subterraneus*)	95.2% (*B. lapidarius*)	73.0%	±20.0
B. pascuorum	3.8% (*B. muscorum*)	97.1% (*B. barbutellus)*	72.2%	±28.1
B. muscorum	3.8% (*B. pascuorum*)	97.1% (*B. bohemicus*)	71.6%	±27.9
B. sylvarum	25.0% (*B. subterraneus*	96.2% (*B. hypnorum*)	71.4%	±19.4
B. humilis	3.8% (*B. muscorum*)	94.2% (*B. barbutellus*)	70.7%	±26.8
B. pratorum	40.4% (*B. sylvarum*)	96.2% (*B. barbutellus)*	70.5%	±16.9
B. hortorum	9.6% (*B. jonellus*)	94.2% (*B. rupestris*)	68.0%	±23.9
B. jonellus	9.6% (*B. ruderarius*)	95.2% (*B. hypnorum*)	67.8%	±23.5
B. soroensis	26.0% (*B. hortorum*)	96.2% (*B. rupestris*)	67.3%	±20.7
B. distinguendus	28.8% (*B. pascuorum*)	97.1% (*B. hypnorum*)	66.7%	±22.6
B. subterraneus	25.0% (*B. sylvarum*)	92.3% (*B. hypnorum*)	63.4%	±22.7

As the results for mismatch images was highly variable, we also analysed this data by 'tail' colour (Fig. 3.2). Misidentification in intra-colour grouping was very high in the browns (76% ± SD 18.1), with just 4% of participants recognising that the images of *B. muscorum* and *B. humilis*, and of *B. muscorum* and *B. pascuorum* were showing two different species. For the mismatch of *B. humilis* and *B. pascuorum* only 8% of participants recognised them as different species (Fig. 3.2). Other outliers in Figure 3.2 for these three browns were for images paired with *B. distinguendus*, for which the corresponding score does not appear as an outlier as there is a larger variation of accurate answers than for the other browns. Average misidentification for the reds was lower (28% ± SD 20.4), with two intra-group pairs showing as outliers, namely *B. lapidarius* and *B. ruderarius* (17% accuracy) and *B. lapidarius* and *B. rupestris* (48% accuracy) (Fig. 3.2).

Table 3.2. Accuracy in identifying images of two different species depending on tail colour. Accuracy is higher when species had different tail colour, but notably lower when both species had brown tails.

	Brown	**Red**	**White**
Brown	19%	-	-
Red	74%	67%	-
White	83%	86%	69%

Although whites formed a larger category (*n* = 10), average misidentification was similar to that seen in the reds (27% ± SD 21.1). There was only one inter-colour group outlier, which was *B. sylvarum* (red) paired with *B. subterraneus* (white). Although the identification guide referred to in this study (Falk 2015) did not define a similar species for *B. sylvarum*, and suggested that the species most confused with *B. subterraneus* is *B. ruderatus*, this study found that 68% of participants erroneously thought that *B. sylvarum* and *B. subterraneus* were a match pair (Table 3.3). Finally, it is notable that across all tail colours, the species without stripes showed the least variation in accurate answers, with the three of the brown species having very similar results

3.4.3 Purported confused species

Using species descriptions in a recently published field guide (Falk 2015), *Bombus* species noted as being ‘confused’ with the model species used in this study (queens only) were noted. These most ‘confused’ species were compared with the findings from this matching study (Table 3.3). Confusion was reported by using *inaccurate* answers on mismatch trials, i.e., by disregarding ‘don’t know’ answers (3.6%). Queens found to be similar in this study were the same as those noted in the guide for 9 of the 20 species (Table 3.3). For five of the model bumblebee queens, the guide did not provide a similar species as they were considered ‘distinctive’, or similarities only referred to male castes. However, for three of these five species (*B. distinguendus, B. pratorum* and *B. sylvarum),* over 50% of participants in this study confused them with another species (Table 3.3). One notable difference is the confused species for *B. soroensis:* the guide suggests that the similar species is *B. lucorum*, but only 31% of participants confused these two species, whereas 71% confused it with *B. hortorum*.

Table 3.3. Comparison of 'confused' species listed in identification guide with results from this study. Misidentification rates are the percentage of participants identifying the pair of images as the same species when they are different species. Scores from this task were in agreement 10 of the 18 similar species noted in the identification guide.

Species	Most confused species		Misidentification score	Agreement
	id guide	Study		
B. humilis	*B. muscorum*	*B. muscorum*	95%	Y
B. muscorum	*B. humilis*	*B. humilis*	95%	Y
B. pascuorum	*B. muscorum*	*B. muscorum*	94%	Y
B. hortorum	*B. jonellus*	*B. jonellus*	89%	Y
B. jonellus	*B. hortorum*	*B. hortorum*	89%	Y
B. pascuorum	*B. humilis*	*B. humilis*	89%	Y
B. lapidarius	*B. ruderarius*	*B. ruderarius*	81%	Y
B. ruderarius	*B. lapidarius*	*B. lapidarius*	81%	Y
B. soroensis	*B. lucorum*	*B. hortorum*	71%	N
B. subterraneus	*B. ruderatus*	*B. sylvarum*	68%	N
B. sylvarum	None	*B. subterraneus*	68%	n/a
B. distinguendis	None	*B. pascuorum*	63%	n/a
B. ruderatus	*B. subterraneus*	*B. subterraneus*	62%	Y
	B. hortorum	*B. hortorum*	57%	Y
B. pratorum	None	*B. sylvarum*	54%	n/a
B. monticola	None	*B. distinguendis*	49%	n/a
B. rupestris	*B. hortorum*	*B. lapidarius*	48%	N
B. barbutellus	*B. sylvestris*	*B. ruderatus*	39%	N
B. terrestris	*B. lucorum*	*B. pratorum*	32%	N
B. lucorum	*B. terrestris*	*B. soroensis*	31%	N
B. bohemicus	*B. vestalis*	*B. barbutellus*	24%	N
B. rupestris	*B. subterraneus*	*B. sylvarum*	22%	N
B. hypnorum	None	*B. barbutellus*	10%	n/a
		B. bohemicus	10%	n/a
		B. soroensis	10%	n/a

Finally, we focused on the cuckoo bumblebees (subgenus *Psithyrus*) and their hosts: *B. barbutellus* (hosts *B. hortorum* and *B. ruderatus*), *B. bohemicus* (host *B. lucorum*) and *B. rupestris* (host *B. lapidarius*), which are of interest in this context as they mimic their hosts. Of these three cuckoo species, participants misidentified two of them with their hosts, *B. rupestris* (48%) and *B. barbutellus* (39%) (Table 3.3). For the other cuckoo and host combinations, a lower number of participants misidentified them as the same species, *B. barbutellus* and *B. hortorum* (25%) and *B. bohemicus* and *B. lucorum* (20%). These results reinforce the lack of continuity or direction when visually identifying inter-species variation.

3.5 Discussion

This study investigated to what extent non-specialists were able to match images of species unfamiliar to them, and if there were any characteristics that influenced whether particular species combinations were more easily misidentified than others. Overall performance was highly variable, with large differences in performance between individual participants, and variation in accuracy depending on which two species were compared in the task.

The majority of participants had no experience of bumblebee identification and only 9% had used an identification guide specifically for bumblebees, but accuracy based on visual cues was varied across these novice participants. For match pairs, average participant accuracy ranged from 70-100%. Mismatch pairs showed a mean accuracy at 76% (Fig. 3.1), meaning that overall 1-in-4 images were mistakenly identified as being the same species when they were, in fact, different. Moreover, individual performance varied greatly, with participant accuracy ranging from 11 to 100%. These findings show a large fluctuation in individual ability. Novice observers achieving 100% accuracy in matching demonstrates the importance of visual perception skills in the process of species identification. Specialist knowledge is crucial in any domain, but the lack of such knowledge does not equate to an inability to identify a species on solely visual cues, and sometimes non-experts are able to outperform experts when matching images (Austen, Bindemann and Roberts 2016). The one person that reported experience 'beyond common' species scored 89% (Fig. 3.1).

When analysed by image, only two of the 20 match pairs were correctly identified as the same species by *all* participants. For mismatch images, accuracy ranged from 4 to 98% across 190 images, showing that intra and inter species variation is vast and unequal. Even though guides provide assistance in the identification process by suggesting similar or 'confusable' species, these suggested species were different to approximately half of the species that this study found to be the most similar (Table 3.3). Many identification guides group images according to tail colour, but as Table 3.2 shows, this is not uniform across species either.

Misidentification within tail colour groups was significantly higher in the browns (76%), compared to the reds (28%) and whites (27%). Three of the browns (*B. humilis*, *B. muscorum* and *B. pascuorum*) were readily mistaken for each other (4-8% accuracy), as well as some confusion with the other brown *B. distinguendus* (29-37% accuracy) One notable point of this confusion is that the three highly confused browns all have the same colour pattern of a block of brown hair on the thorax and abdomen, but no stripes. The identification guide did not identify a 'confused' species for *B. distinguendus*, and although there is a high variation in accuracy for images containing *B. distinguendus*, 63% of participants confused it with *B. pascuorum* (Table 3.3). These results show *B. pascuorum* to be readily confused with all other browns, yet a recent paper comparing biological records from skilled naturalist with lay citizen science records chose *B. pascuorum* as a target species, as it is "widespread and fairly easy to identify" (Wal, Anderson and Robinson 2015). The authors chose *B.*

pascuorum and *B. hypnorum* as focal species due to them being the two species that were recorded most often in the datasets, and although the latter was the least confused in this study, ease of identification is subjective and based on the judgement of the authors, who are likely to be familiar with the species. The reds with least variation in accuracy were *B. lapidarius*, *B. ruderarius* and *B. rupestris*, which apart from their tail colouration, have no stripes or other colouration. As with the browns, the highest rates of misidentification occurred between these three non-striped, red-tailed species, with 17% of participants accurately recognising that *B. lapidarius* and *B. ruderarius* were different species, and 48% that *B. lapidarius* and *B. rupestris* are not the same. Accuracy in telling *B. ruderarius* and *B. rupestris* apart was higher at 75%, but when compared side-by-side, the difference in their shapes are noticeable.

Cuckoo bumblebees present an interesting group to investigate. As obligate parasites they are not found outside the ranges of their host species, and in Europe share analogous colour patterns significantly more than would be expected by chance (Reinig 1935; Williams 2008). Although evidence shows that cuckoos can use chemical mechanisms to gain entries to hosts' nests (Dronnet et al. 2005; Martin et al. 2010), there is wide agreement that similarity in colour patterns with hosts is due to Müllerian mimicry (Martin et al. 2010; Plowright and Owen 1980). This mutual resemblance of warning colouration relies on conspicuous markings to deter predators that may have encountered similar looking, but harmful, organisms in the past. Such deception may account for misidentification by predators, but while it is based primarily on colours and patterns, human vision processes a substantial amount of information based on shape or form as well (Du, Wang and Zhang 2007). Cuckoo bees lack the pollen baskets found on the legs of true bees, but knowledge of such distinguishing features would depend on experience (Baruch, Kimchi and Goldsmith

2014), or information provided in training material, and a novice in bumblebee identification is unlikely to know this detail. Although two of the three cuckoo species had the highest misidentification scores with their hosts (Table 3.3), all of these scores were less than 50%. While cuckoo and host species ranges naturally overlap, the consequences of misidentification will not only be a false representation of species numbers, but their differing behaviours mean that the roles they play in the ecosystem are not equal (Cardinal et al. 2010; Dronnet et al. 2005)

Beyond picture matching, a guide contains species information that may support, or discount, the accuracy of the observation. Mayr (1966) suggests that despite being visually similar, some potentially confusable species should be easily discernible due to their life histories. Such distinctions may be achievable by those with prior knowledge of the observed species, but a novice relying on an identification guide to name an organism, the process is quite different. Identification guides provide information on species' behaviours or distributions but, the observer is required to provisionally identify and name a species *before* they can refer to this information. Moreover, some, descriptions pinpoint distinguishing characteristics, not easily applied in the field. For example, *B. hortorum and B. jonellus,* which were thought to be the same species by 89% of participants in this study differ in size (6 mm compared to 10 mm). This information on size may aid field identification if there is a reference point but, even then, all castes can exhibit dwarfism (Falk 2015), and identification accuracy can diminish over distance (Lampinen et al. 2014). Furthermore, hard copy field guides provide knowledge at a static point in time, whereas numbers and distributions are dynamic (Wiederholt et al. 2015) (e.g. extinction, abundance, distribution), even over short periods of time (Tayleur et al. 2016).

This study found the most similar species to differ to those suggested in the identification guide (Table3.3). The reasons underpinning an author's decision to define which species are confused with each other is subjective, but is likely to be influenced by prior knowledge and experience. For example, while some species are visually similar, they may not occur in the same geographic area, which could influence this decision. Again, this background knowledge is unlikely to be known by a non-specialist, plus, distributions can change and range expansion or reduction may occur quicker than a new guide is published. For example, Falk (2015) provides distribution maps based on data from the Bees Wasps and Ants Recording Society (BWARS), dated 2000 onwards. Although believed to be extinct in the UK from the late 1980s, a description of the short-haired bumble bee (*Bombus subterraneus*) is included in this guides, as there has been a 'reintroduction into Southeast England started in 2015' (pg. 405). The reintroduction programme is monitored by the Bumblebee Conservation Trust (BBCT) in Dungeness, who show post-2000 records on their website (www.bumblebeeconservation.org), and notes that this BWARS data is yet to be added. This example also highlights the time delay of some observations being digitised, and that publicly accessible datasets are not always comprehensive. Furthermore, differences are not necessarily uniform in a global context. For example, the aforementioned cuckoo resemblance to host being higher than chance in Europe, but not in North America (Williams 2008). Assessing the similarity of two things is an important part of cognition, but judgements can vary depending on the comparisons being made (Farell 1985; Goldstone, Day and Son 2010), or the subject itself (Palmer and Schloss 2010). There are occasions when specimens are collected in such large volumes that researchers recruit local people to perform an initial sort into groups based on morphology (parataxonomy, see Krell (2004)). However, interpretation of shape and colour can differ between cultures (Roberson and Davidoff 2006a; Kay and

Regier 2007), and if projects are being conducted in different communities translation may include changes in more than just text.

Recognising where errors occur in species identification could mitigate against incorrect population estimates (Runge, Hines and Nichols 2007; Dickinson, Zuckerberg and Bonter 2010; Shea et al. 2011), aid future monitoring (Murphy 1989; Guisan et al. 2013), and be considered in studies on extinction (Roberts, Elphick and Reed 2010). In practice, novices recorders participating in citizen science projects will receive training and feedback, which is crucial to learning (Tulloch et al. 2013), but there are numerous examples where training and experience is unknown, for example with occurrence data (Pocock et al. 2015; Todd et al. 2016) and museum collections (Reinig 1935; Goodwin et al. 2015). Furthermore, identifying which species are easily confused can help when choosing which model species to monitor. Given the extensive literature that exists in psychology relating to the mechanisms involved in this study, we suggest that species identification can be aided by investigating how guides are used, as well as their content. For example, perceptions and expectancy can affect decisions in forensic face matching (Kassin et al. 2013), and decisions on species identification may be influenced by information in a guide, such as the likelihood of that species occurring in particular geographic regions. Well delivered citizen science projects produce visual aids designed to meet the aims of that project, but with generic biological recording, the requirements of a guide for a specific recording scheme aimed at a certain audience (e.g. schoolchildren, novice observers) will differ from that purchased as a reference for an enthusiast. The results from this study suggests that identification is an individual ability, and testing participants prior to any training could be a benchmark by which to evaluate the training, or tailor it to suit individual needs.

3.6 References

Allen, D. (1976). *The Naturalist in Britain: A Social History*. Princeton University Press, Princeton.

Austen, G.E., Bindemann, M. and Roberts D.L. (2016). Species identification by experts and non-experts: comparing images from field guides. *Scientific Reports* **6**, 33634.

Baruch, O., Kimchi, R. and Goldsmith, M. (2014). Attention to distinguishing features in object recognition. *Visual Cognition*, **22**, 1184-1215.

BBC Wildlife Pocket Guides Number 11 (2006). *Bumblebees other bees and wasps*. BBC Wildlife Magazine, Bristol.

Beerkircher, L., Arocha, F. and Barse, A. (2009). Effects of species misidentification on population assessment of overfished white marlin *Tetrapturus albidus* and roundscale spearfish *T. georgii*. *Endangered Species Research*, **9**, 81-90.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller R., Edwards, M., Peeters, T., Schaffers, A.P., Potts, S.G., Kleukers, R., Thomas, C.D., Settele, J. and Kunin, W.E. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, **313**, 351-354.

Bumblebee Conservation Trust (BBCT) www.bumblebeeconservation.org/white-tailed-bumblebees/short-haired-bumblebee/ [Accessed 3rd March 2018].

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., et al. (2010). Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Cameron, S.A., Hines, H.M. and Williams, P.H. (2006). A comprehensive phylogeny of the bumble bees (*Bombus*). *Biological Journal of the Linnean Society*, **91,** 161–188.

Cardinal, S., Straka, J. and Danforth, B.N. (2010). Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *PNAS*, **37**, 16207-16211.

Casey, L.M., Rebelo, H., Rotheray, E. and Goulson, D. (2015). Evidence for habitat and climatic specializations driving the long-term distribution trends of UK and Irish bumblebees. *Diversity and Distributions*, **21**, 864–875.

Cohn, J.P. (2008). Citizen Science: can volunteers do real research? *BioScience*, **58**, 192-197.

Cole, F.J. (1944). *A history of comparative anatomy from Aristotle to the eighteenth century*. Macmillan.

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J. and Waller, D.M. (2011). Assessing citizen science data quality: an invasive species case study. *Conservation Letters* **4,** 433–442.

Curtis, V. (2015). *Online citizen science projects: an exploration of motivation, contribution and participation*. The Open University. PhD thesis.

Dickinson, J.L., Zuckerberg, B. and Bonter, D.N. (2010). Citizen Science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 149-172.

Dronnet, S., Simon, X., Verhaeghe, J.C., Rasmont, P. and Errard, C. (2005). Bumblebee inquilinism in *Bombus* (Fernaldaepsithyrus) sylvestris (Hymenoptera, Apidae): behavioural and chemical analyses of host-parasite interactions. *Apidologie,* **36**, 59–70.

Du, J-X., Wang, X-F. and Zhang, G-J. (2007). Leaf shape based plant species recognition. *Applied Mathematics and Computation*, **185**, 883–893.

Duelli, P. (1997). Biodiversity evaluation in agricultural landscapes: an approach at two different scales. *Agriculture, Ecosystems & Environment,* **62**, 81–91.

Dunlap, T. (2005). Tom Dunlap on early bird guides. *Environmental History*, **10**, 110-118.

Edwards, M. and Jenner, M. (2005). *Field guide to the bumblebees of Great Britain & Ireland*. Ocelli Ltd, Eastbourne.

Ellwood, E., Bart, Jr., H.L., Doosey, M.H., Jue, D.K., Mann, J.G., Nelson, G., Rios, N. and Mast, A.R. (2016). Mapping Life – quality assessment of novice vs. expert georeferencers. *Citizen Science: Theory and Practice*, **1**, 1-12.

Elphick, C.S. (2008). How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, **45**, 1313–1320.

Falk, S.J. and Lewington, R. (2015). *Field guide to the bees of Great Britain and Ireland*. Bloomsbury, London.

Farell, B. (1985). ‘Same’–‘different’ judgments: a review of current controversies in perceptual comparisons. *Psychological Bulletin*, **98**, 419-456.

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J. Stevenson, R.D., Courtney, G.W., Van Dyk, J.K. and Ellison A.M. (2013). Next-generation field guides. *BioScience*, **63**, 891–899.

Fitzpatrick, M.C., Preisser, E. L., Ellison, A. M. and Elkinton, J. S (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, **19**, 1673–1679.

Fore, L.S., Paulsen, K. and O’Laughlin, K. (2001). Assessing the performance of volunteers in monitoring streams. *Freshwater Biology*, **46**, 109–123.

Gardiner, M.M., Allee, L.L., Brown, P.M., Losey, J.E., Roy, H.E. and Smyth, R.R. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Frontiers in Ecology and the Environment,* **10**, 471–476.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ,* **3**, e1303.

Givens, J., Reeds, K. and Touwaide, A. (2006). *Visualizing Medieval Medicine and Natural History, 1200-1550* (Vol. 5). Ashgate Publishing Ltd., Aldershot.

Goldstone, R., Day, S. and Son, J. (2010). Comparison. In: Glatzeder, B., Goel, V. and von Müller, A. eds. *Towards a Theory of Thinking*. Springer, Berlin, 103–122.

Gollan, J., de Bruyn, L.L., Reid, N. and Wilkie, L. (2012). Can volunteers collect data that are comparable to professional scientists? A study of variables used in monitoring the outcomes of ecosystem rehabilitation. *Environmental Management,* **50**, 969–978.

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I. and Scotland, R.W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology,* **25**, R1066–R1067.

Goulson, D., Hanley, M., Darvill, B., Ellis, J.S. and Knight, M.E. (2005). Causes of rarity in bumblebees. *Biological Conservation,* **122**, 1–8.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R. Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield,S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. and Buckley, Y.M. (2013). Predicting species distributions for conservation decisions. *Ecology letters*, **16**, 1424–1435.

Joppa, L.N., O'Connor, B., Visconti, P., Smith, C., Geldmann, J., Hoffmann, M., Watson, J.E., Butchart, S.H., Virah-Sawmy, M., Halpern, B.S. and Ahmed, S.E. (2016). Filling in biodiversity threat gaps. *Science*, **352,** 416–418.

Juffe-Bignoli, D., Brooks, T.M., Butchart, S.H., Jenkins, R.B., Boe, K., Hoffmann, M. et al. (2016). Assessing the Cost of Global Biodiversity and Conservation Knowledge. *PLoS One*, **11**, e0160640.

Kassin, S.M., Dror, I.E. and Kukucka, J. (2013). The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, **2**, 42–52.

Kay, P. and Regier, T. (2007). Color naming universals: the case of Berinmo. *Cognition*, **102**, 289-298.

Krell, F.T. (2004). Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of ‘morphospecies’ sorting. *Biodiversity and Conservation*, **13**, 795–812.

Kremen, C., Ullman, K.S. and Thorp, R.W. (2011). Evaluating the quality of citizen-scientist data on pollinator communities. *Conservation biology : the journal of the Society for Conservation Biology*, **25**, 607–617.

Lampinen, J.M., Erickson, W.B., Moore, K.N. and Hittson, A. (2014). Effects of distance on face recognition: implications for eyewitness identification. *Psychonomic Bulletin & Review*, **21**, 1489–1494.

Lawrence, A. (2009). The first cuckoo in winter: phenology, recording, credibility and meaning in Britain. *Global Environmental Change*, **19**, 173-179.

Martin, S.J., Carruthers, J.M., Williams, P.H. and Drijfhout, F.P. (2010). Host specific social parasites (*Psithyrus*) indicate chemical recognition system in bumblebees. *Journal of Chemical Ecology*, **36**, 855–863.

Mayr, E. (1966). *Animal Species and Evolution*. Harvard University Press, Cambridge.

Merenlender, A., Crall, A., Drill, S., Prysby, M. and Ballard, H. (2016). Evaluating environmental education, citizen science, and stewardship through naturalist programs. *Conservation Biology*, **30**, 1255-1265.

Miller-Rushing, A., Primack, R. and Bonney, R. (2012). The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, **10**, 285-290.

Murphy, D.D. (1989). Conservation and confusion: wrong species, wrong scale, wrong conclusions. *Conservation Biology*, **3**, 82-84.

Özbek, M. and Bindemann, M. (2011). Exploring the time course of face matching: Temporal constraints impair unfamiliar face identification under temporally unconstrained viewing. *Vision Research*, **19**, 2145-2155.

Palmer, S.E. and Schloss, K.B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, **107**, 8877–8882.

Paxton, R. (2015). Entomology: The bee-all and end-all. *Nature*, **521**, S57-59.

Pinchen, B.J. (2010). *A pocket guide to the bumblebees of Britain and Ireland*, Forficula Books, Lymington.

Plowright, R.C. and Owen, R.E. (1980). The evolutionary significance of bumble bee colour patterns: a mimetic interpretation. *Evolution*, **34**, 622-637.

Pocock, M.J.O., Roy, H.E., Preston, C.D. and Roy, D.B. (2015). The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnean Society*, **115**, 475–493.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W.E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution*, **25,** 345-353.

Prŷs-Jones, O.E. and Corbet, S.A. (1991) *Bumblebees*. Richmond Publishing Company Limited, London.

Reinig, W.F. (1935). On the variation of *Bombus lapidarius* L. and its cuckoo, *Psithyrus rupestris* Fabr., with notes on mimetic similarity. *Journal of Genetics*, **30**, 321–356.

Roberson, D. and Davidoff, J. (2006). Colour categories and category acquisition in Himba and English. In: Pitchford, N.J. and Biggam, C.P. eds. *Progress in Colour Studies* (Vol. 2). John Benjamins Publishing Co., Amsterdam, 159-172.

Roberts, D.L., Elphick, C.S. and Reed, J.M. (2010). Identifying anomalous reports of putatively extinct species and why it matters. *Conservation biology*, **24**, 189–196.

Robinson, B.S., Inger, R. and Gaston, K.J. (2016). A rose by any other name: plant identification knowledge & socio-demographics. *PLoS One*, **11**, e0156572.

Rodrigues, A.S.L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M. and Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in Ecology and Evolution*, **21**, 71–76.

Runge, J., Hines, J. and Nichols, J. (2007). Estimating species-specific survival and movement when species identification is uncertain. *Ecology*, **88**, 282-288.

Scharf, S.T. (2009). Identification keys, the ‘natural method,’ and the development of plant identification manuals. *Journal of the History of Biology*, **42**, 73–117.

Scriven, J.J., Woodall, L. C., Tinsley, M. C., Knight, M. E., Williams, P. H., Carolan, J. C., Brown, M. J. F. and Goulson, D. (2015). Revealing the hidden niches of cryptic bumblebees in Great Britain: implications for conservation. *Biological Conservation*, **182**, 126–133.

Shea, C.P., Peterson, J.T., Wisniewski, J.M. and Johnson, N.A. (2011). Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.

Stevenson, R., Haber, W. and Morris, R. (2003). Electronic field guides and user communities in the eco-informatics revolution. *Conservation Ecology*, **7**, 3.

Sutherland, W.J., Roy, D.B. and Amano, T. (2015). An agenda for the future of biological recording for ecological monitoring and citizen science. *Biological Journal of the Linnean Society*, **115**, 779–784.

Tayleur, C.M., Devictor, V., Gaüzère, P., Jonzén, N., Smith, H.G. and Lindström, Å. (2016). Regional variation in climate change winners and losers highlights the rapid loss of cold-dwelling species. *Diversity and Distributions*, **22**, 468-480.

Tillett, B.J., Field, I.C., Bradshaw, C.J., Johnson, G., Buckworth, R.C., Meekan, M.G. and Ovenden, J.R. (2012). Accuracy of species identification by fisheries observers in a north Australian shark fishery. *Fisheries Research*, **127**, 109–115.

Todd, B.D., Rose, J.P., Price, S.J. and Dorcas, M.E. (2016). Using citizen science data to identify the sensitivity of species to human land use. *Conservation Biology*, **30**, 1266-1276.

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J. and Martin, T.G. (2013). Realising the full potential of citizen science monitoring programs. *Biological Conservation*, **165**, 128-138.

Van der Wal, R., Anderson, H. and Robinson, A., Sharma, N., Mellish, C., Roberts, S., Darvill, B. and Siddharthan, A. (2015). Mapping species distributions: a comparison of skilled naturalist and lay citizen science recording. *Ambio*, **44**, 584-600.

Wiederholt, R., Trainor, A.M., Michel, N., Shirey, P.D., Swaisgood, R.R., Tallamy, D. and Cook-Patton, S.C. (2015). The face of conservation responding to a dynamically changing world. *Integrative Zoology*, **10**, 436-452.

Williams, P. (2007). The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. *Biological Journal of the Linnean Society*, **92**, 97-118.

Williams, P.H. (2008). Do the parasitic *Psithyrus* resemble their host bumblebees in colour pattern? *Apidologie*, **39**, 637–649.

Williams, P.H. and Osborne, J.L. (2009). Bumblebee vulnerability and conservation world-wide. *Apidologie*, **40**, 367-387.

Zapponi, L., Cini, A., Bardiani, M., Hardersen, S., Maura, M., Maurizi, E., De Zan, L.R., Audisio, P., Bologna, M.A., Carpaneto, G.M. and Roversi, P.F., Sabbatini Peverieri, G., Mason, F. and Campanaro, A. (2016). Citizen science data as an efficient tool for mapping protected saproxylic beetles. *Biological Conservation*, **208**, 139-145.

Dog or mop? ©teenybiscuit

Chapter 4.

4 Species identification by conservation practitioners using online images: accuracy and agreement between experts

Revised following peer review from *Peer J* (15838)

4.1 Abstract

Emerging technologies have led to an increase in species observations being recorded via digital images. Such visual records are easily shared, and are often uploaded to online communities when help is required to identify or validate species. Although this is common practice, little is known about the accuracy of species identification from such images. Using online images of newts that are native and non-native to the UK, this study asked holders of great crested newt (*Triturus cristatus*) licences (issued by UK authorities to permit surveying for this species) to sort these images into groups, and to assign species names to those groups. All of these experts identified the native species, but agreement among these participants was low, with some being cautious in committing to definitive identifications. Individuals' accuracy was also independent of both their experience and self-assessed ability. Furthermore, mean accuracy was not uniform across species (69-96%). These findings demonstrate the difficulty of accurate identification of newts from a single image, and that expert judgements are variable, even within the same knowledgeable community. We suggest that identification decisions should be made on multiple images, and verified by more than one expert, could improve the reliability of species data.

4.2 Introduction

The increasing reliability of new technologies has enabled those interested in the natural world to observe, identify and count species in a faster, cheaper and less intrusive manner than ever before (Pimm et al., 2015). One such use of these tools is electronic image capture from smart phones, camera traps, videos and drone footage. These images can be used to identify species (Cooper, Shirk, & Zuckerberg, 2014; Daume & Galaz, 2016; McKinley et al., 2016; O'Donnell & Durso, 2014; Pocock, 2016), and are often accompanied by informative metadata (for example, date, time and location), thus providing a wealth of information regarding species numbers, distributions and behaviours. Furthermore, these images permit identification and validation to take place at a later date, and can be shared relatively easily. A good example of this is the submission of photographs from people who wish to identify a species or validate their observation. Whether as part of a citizen science project, through a local recording group, or simply ad-hoc observations, the process involves images being uploaded electronically for identification or verification by enthusiasts of varying expertise, for example, using iSpot (www.ispotnature.org), iNaturalist (www.inaturalist.org), iRecord (www.brc.ac.uk/irecord/) and reddit (www.reddit.com/r/species) (Bates et al., 2015; Silvertown et al., 2015; Burgess et al., 2016; Daume & Galaz, 2016; Leighton et al., 2016). These online communities conduct and collaborate in species identification, but cannot always do so with certainty. For example, in iSpot contributors make a 'Likely ID' to remind participants that identification from images lacks certainty (Silvertown et al., 2015). Expert judgement in the identification of specimens can be sought through other means, such as wildlife trusts, local recording groups and county recorders. However, in this instance, it may be that only one person identifies or verifies the image for recording purposes. If this is then referred to another specialist and that person disagrees with this identification, the observation may be recorded to an agreed taxonomic level (e.g. genus). Irrespective of the route taken, expert identification and validation is widely sought.

In this context, deciding who is an expert and how expert judgements can be verified remain open questions (Goldman, 2001; Burgman et al., 2011). Society turns to individuals with certain skills and experience for advice in decision-making (Burgman, 2015). However, this experience is often linked to qualifications or perceived ability rather than validated performance (Farrington-Derby and Wilson 2006). Moreover, although expert knowledge can be generalised as information about a subject that is not universally known (Martin et al. 2012), it is often difficult to define (Hoffman, 1996). In general, expertise is domain dependent (Hoffman, 1996; Chi, 2006), dynamic (Lave & Wenger, 1991), influenced by social status (Stebbins, 1977; Ericsson, 2014), and unequally distributed within communities (Evans, 2008). Experience can make processes more automatic and reduce the effort required to complete a task, but does not necessarily lead to improved performance (Hoffman, Crandall, & Shadbolt, 1998; Ericsson, 2014; Austen et al., 2016). However, defining who is an expert can be subjective and contentious, and may be perceived differently by those within a community to those outside of it (Lave & Wenger, 1991; Goldman, 2001; Ericsson, 2014; Burgman, 2015). Indeed, even within specialist communities, experts are likely to recognise certain individuals as more competent than others, and have a perception of how their own expertise compares with the rest of their peers (Hoffman 1996). Ultimately, however, if an individual is perceived as an expert, they will be asked for their advice or judgement.

One area of conservation practice that relies on expert judgement is ecological monitoring (Burgman et al., 2011; Kapos et al., 2009; Lindenmayer & Likens, 2010). These data provide information on species numbers and distributions, including species of conservation concern, and invasive taxa that pose a threat to those species (Farnsworth et al., 2013; Latombe et al., 2016; Mang et al., 2016). One type of monitoring that incorporates both protected and invasive species is great crested newt (*Triturus cristatus*) monitoring in the UK. As a European Protected Species (EPS), *T. cristatus* is protected under law and anyone planning to survey or handle this species requires a licence. Whether professionals or volunteers, applicants for EPS licences may be expected to be familiar with native newt species. However, instead of being required to demonstrate the relevant identification skills, applicants are required supply a written reference from another licence holder (see www.gov.uk). Therefore, the issue of licences is influenced by the subjective opinion of other experts within that community, rather than based on an objective demonstration of a specific level of competence to the issuing authorities.

This is an important issue because errors in identification can prove costly when surveying for newts. In addition to three native newt species, the UK is also home to a number of non-native newts, of which the alpine newt (*Ichthyosaurus alpestris*) is the most widespread. This invasive species is capable of carrying the lethal amphibian fungal pathogens *Batrachochytrium dendrobatidis* (Ohst, Gräser, & Plötner, 2013) and *B. salamandrivorans* (Spitzen-van der Sluijs, 2016). If this species is caught in error, it cannot be released into the wild, as this is illegal under the Wildlife and Countryside Act (1981). Error in newt identification can also lead to delays in development, unnecessary mitigation, fines for breaching the terms of a licence, or potential spread of pathogens. For this reason, misidentification of newts can prove costly. For example, the annual cost of great crested newt mitigation in England alone is estimated to be between £20 million and £43 million (Lewis, Griffiths, & Wilkinson 2017).

In this study, we sought to investigate the likelihood of errors in newt identification. For this purpose, we invited holders of great crested newt licences, which allows individuals to survey this species in accordance with the European Conservation of Habitats and Species Regulations 2010 and the UK Wildlife and Countryside Act 1981, to perform a simple image-sorting task. Internet images were selected to investigate how experts group and name images of different newt species. We also explored whether this was linked to (i) self-assessed identification ability, (ii) perceived identification ability in comparison with peers, and (iii) experience. We also compared accuracy between professional and volunteer surveyors. Finally, we investigated whether the presence of certain diagnostic characteristics in these images were linked to increased levels of identification.

4.3 Methods

4.3.1 Image sorting task

This research was approved by the Ethics Committee of the School of Anthropology and Conservation at the University of Kent, and conducted in accordance with the ethical guidelines of the British Psychological Society.

At the UK's annual Herpetofauna Workers Meeting (HWM) in 2015, individuals with a great crested newt licence were invited to participate in a photo-sorting task. Seventeen participants (15 male, 2 female, mean age = 43 years ± 13) completed the task both at the event and later at the University of Kent. All reported good or corrected-to-normal vision with glasses or contact lenses. Informed consent was obtained from participants.

Four species of newt found in the UK were chosen as stimuli, three native and one non-native. Native species were the smooth newt (*Lissotriton vulgaris*), palmate newt (*L. helveticus*) and great crested newt (*T. cristatus*), and the non-native species was the alpine newt (*I. alpestris*). These four species are found throughout the UK and EPS licence holders are likely to come into contact with when surveying for *T.cristatus* (nbnatlas.org). Using the Latin binominal, photographs for each species were retrieved via Google© search engine under the 'images' option. Twenty unique images were chosen for each of the four species, from various websites (Appendix 4.1). Image selection was aimed at incorporating the range of variability that may be encountered by observers in the field. For example, selected images included males and females, newts in terrestrial and aquatic stages, and with from various perspectives (i.e. dorsal, ventral, lateral and part views). The label from the downloaded image was taken as the correct species description. Although the species name assigned to the image may be incorrect, the websites of recognised organisations (e.g. conservation non-government organisations, herpetofauna fora, Wildlife Trusts, educational websites, national news outlets, etc.) took preference. The specialist nature of these sites suggests an element of validation before labelling the images, although no sources are guaranteed to be error-free. Images were randomised and numbered 1 to 80. These numbers were used to mark the reverse of the photographs used in the sorting task. This approach is used in other research areas, such as the study of forensic human face recognition (see Jenkins et al., 2011), and provides a highly-controlled scenario for newt identification. This study design also eliminates other non-visual factors (e.g. where and when the images were taken), which removes bias associated with prior knowledge of breeding cycles and species distribution.

Participants were asked to complete a short questionnaire, including age, gender and their experience with surveying amphibians in the UK (Appendix 4.2). This included their experience with surveying for the target species, self-perceived identification ability, self-perceived identification ability in relation to their peers, and whether they surveyed in a volunteer or professional capacity. Categories for surveying included professional, volunteering with local recording groups, organised projects, and those who survey independent of any affiliation, such as enthusiasts that monitor an area or species for personal interest. Participants performed a simple sorting task, for which they were asked to sort 80 newt images into piles according to species, irrelevant of gender. Participants were supervised during the task, but no further instructions were given. Participants were also encouraged not to discuss their findings with other volunteers that had yet to participate in the task. No restrictions were placed on the number of piles created, or on time taken, to avoid any undue pressure to complete the task. Once all 80 images were sorted, participants were asked to assign a species name to each group. The images were shuffled for each participant prior to the task.

4.3.2 Diagnostic characteristics

To investigate whether certain aspects of these images influenced identification accuracy, each image was analysed according to the angle of view and which body parts were visible. Scores were made on perspective (namely dorsal, lateral or ventral view), the visibility of diagnostic characteristics and whether the head, head and body, or the whole newt was observable. Despite there being numerous characters defined to aid newt identification, many are linked to breeding condition, especially in males, which creates a gender and temporal bias (Arnold, Burton, & Ovenden, 1978; Arnold & Ovenden, 2002). Therefore, for some defining characters their presence is only indicative for that species at certain times in the breeding cycle. For example, crests are characteristic for breeding *T. cristatus* and *L. vulgaris* males, and webbed hind feet and tail filaments are characteristic in breeding male *L. helveticus*. However, the absence of these characters does not necessarily indicate a different species, but possibly a non-breeding male or a female. Also, other characters such as colouration or 'belly' spots are only visible from certain angles. Furthermore, although the 'warty' skin of *T. cristatus* is a defining feature, it is difficult to code for as not only will skin be visible on all photos, but warty skin will only determine whether that newt is *T. cristatus,* rather than differentiate between the other species. With these considerations, a score of one was allocated if hind feet and the whole of the tail were visible in the images, and zero if not. Whether the animal was photographed in an aquatic or terrestrial situation may have been a factor of interest, but this could not be ascertained from every image and was therefore discounted.

4.3.3 Analyses

Participant accuracy was analysed by self-assessed ability, ability compared with peers, surveying experience and the capacity in which they surveyed. A boxplot comparing accuracy across target species was constructed in R (version 3.4.2). ANOVAs and post-hoc Tukey tests compared grouping of species, and the accuracy depending on how species appeared in photos. A Pearson product-moment correlation was run to determine correlation between participants' experience and accuracy. Inferential statistics were performed in SPSS (IBM SPSS Statistics for Macintosh, Version 24.0).

4.4 Results

Participant experience in newt surveying averaged 13.9 years (range 4-26 years, SD = ±7.6). This was also reflected in how participants rated their identification abilities, which was either ‘very good’ (*n* = 10) or ‘good’ (*n* = 7) on a five-point scale. Moreover, most participants perceived their identification skills as ‘better than’ (*n* = 7) or the ‘same as’ (*n* = 9) their peers on another five-point scale, with only one participant considering themselves to be ‘worse than’ their peers (*n* = 1).

Overall identification accuracy was determined by the species named by participants matching the species named in the downloaded image. In the sorting task, participants created an average of 4.7 (range 4-8, SD = ±1.1) groups of images for the four newt species. Nine of the 17 participants correctly sorted the images into four groups. These participants also assigned the correct names of each of the study species to their groups. However, none of these groups were sorted in the same way by participants, or fully agreed with how images had been labelled online (Table 4.1). A further seven participants assigned the names of the study species to some of their groups, but also created and named further groups. These additional groups were the Italian crested newt (*T. carnifex*) (*n* = 2), palmate/smooth newt hybrid (*n* = 2), palmate *or* smooth newt (*Lissotriton* spp.) (*n* = 1), and unknown (*n* = 5) (Table 4.1). The remaining participant created four groups, with three named as the native newts, and declared the fourth group as ‘unknown’ (Table 4.1)

Table 4.1 Summary of how participants sorted images into groups and the names assigned to those groups. All but one participant recognised the four study species (denoted by *), and eight participants also assigned other names or nominated the group as ‘don’t know’.

Participant	***I. alpestris**** (*n*=20)	***L. helveticus**** (*n*=20)	***L. vulgaris**** (*n*=20)	***T. cristatus**** (*n*=20)	***T. carnifex***	***L. helveticus/L. vulgaris* hybrid**	***Lissotriton* spp.**	**Don’t know**
1	26.3%	21.3%	27.5%	25.0%	-	-	-	-
2	25.0%	18.8%	31.3%	25.0%	-	-	-	-
3	21.3%	17.5%	31.3%	27.5%	-	-	-	2.5%
4	25.0%	16.3%	23.8%	22.5%	1.3%	3.8%	7.5%	-
5	21.3%	16.3%	26.3%	27.5%	-	-	-	8.8%
6	25.0%	28.8%	15.0%	18.8%	-	6.3%	-	6.3%
7	23.8%	20.0%	30.0%	26.3%	-	-	-	-
8	25.0%	25.0%	25.0%	25.0%	-	-	-	-
9	22.5%	21.3%	28.8%	27.5%	-	-	-	-
10	21.3%	25.0%	25.0%	28.8%	-	-	-	-
11	21.3%	25.0%	26.3%	22.5%	5.0%	-	-	-
12	23.8%	22.5%	27.5%	26.3%	-	-	-	-
13	25.0%	25.0%	25.0%	25.0%	-	-	-	-
14	25.0%	21.3%	28.8%	25.0%	-	-	-	-
15	22.5%	17.5%	32.5%	23.8%	-	-	-	3.8%
16	-	10.0%	25.0%	23.8%	-	-	-	41.3%
17	22.5%	17.5%	20.0%	25.0%	-	-	-	15.0%

Mean overall identification accuracy across participants was 82.7% (range 43.8-93.8%, SD = ±12.4). However, when calculating identification accuracy within the four study species only (i.e. discounting unknowns and false positives) the mean increased to 87.2% (range 56.2-95.1%, SD = ±9.8) (Table 4.2). Participants that grouped and named images as just the four study species averaged 90.6% accuracy, whereas those participants that created a 'don't know' pile (*n* = 6) averaged 62.1% overall. However, mean accuracy for these participants increased to 70.1% in the study species only (Table 4.2). As well as variation *between* individual performance, differences in self-perceived ability were found in those with similar scores. For example, the highest overall accuracy score (93.8%) was achieved by three participants, but while two of these participants considered their identification skills to be 'very good' and 'better than' their peers, the third considered themselves to be 'good' and 'same as' their peers (Table 4.2). Furthermore, when measuring accuracy on the study species alone, the highest accuracy was 95.1% by a participant who ranked themself as 'good' yet 'worse than peers' (Table 4.2).

Table 4.2. Participants' self-assessed ability of their own identification skills, self-assessed ability compared to their peers, experience, and accuracy scores. Mean accuracy is reported for all images (overall), and for groups named as study species. The table is ranked (descending order) by own ability, ability compared with peers, then accuracy.

Participant	**Own ability (self-assessed)**	**Ability v Peers (self-assessed)**	**Mean accuracy**		**Experience (years)**
			Overall	**Study species**	
1	Very good	Better than	93.8%	95.0%	20
2	Very good	Better than	93.8%	93.9%	26
3	Very good	Better than	82.5%	86.2%	25
4	Very good	Better than	81.3%	92.8%	12
5	Very good	Better than	80.0%	88.7%	17
6	Very good	Better than	78.8%	91.2%	20
7	Very good	Same as	91.3%	92.0%	6
8	Very good	Same as	90.0%	90.0%	21
9	Very good	Same as	86.3%	86.8%	12
10	Very good	Same as	85.0%	85.5%	20
11	Good	Better than	75.0%	80.3%	4
12	Good	Same as	93.8%	94.0%	11
13	Good	Same as	92.5%	92.5%	25
14	Good	Same as	88.8%	89.1%	6
15	Good	Same as	68.8%	73.1%	7
16	Good	Same as	43.8%	56.2%	7
17	Good	Worse than	81.3%	95.1%	10

Three participants achieved the highest overall accuracy score (93.8%), two of which considered their identification skills to be 'very good' and 'better than' their peers, with the third participant considering themselves to be 'good' but 'same as' their peers (Table 4.2). However, when measuring accuracy on the study species alone, the highest accuracy was 95.1% by a participant who ranked themself as 'good' yet 'worse than peers' (Table 4.2). Conversely, the individual with the lowest score of

43.8%, categorised their identification ability as ‘good’ and ‘same as’ their peers. Their performance for just the study species was higher at 56.2%, but remained poorer than all other participants (Table 4.2). This is the same participant that did not identify that the images contained the non-native alpine newt. A Pearson product-moment correlation of overall accuracy and number of years’ experience with newt surveying did not reach significance ($r = 0.43$, $n = 17$, $p = 0.086$), with mean accuracy on study species alone following the same trend ($r = 0.41$, $n = 17$, $p = 0.104$) (Fig. 4.1).

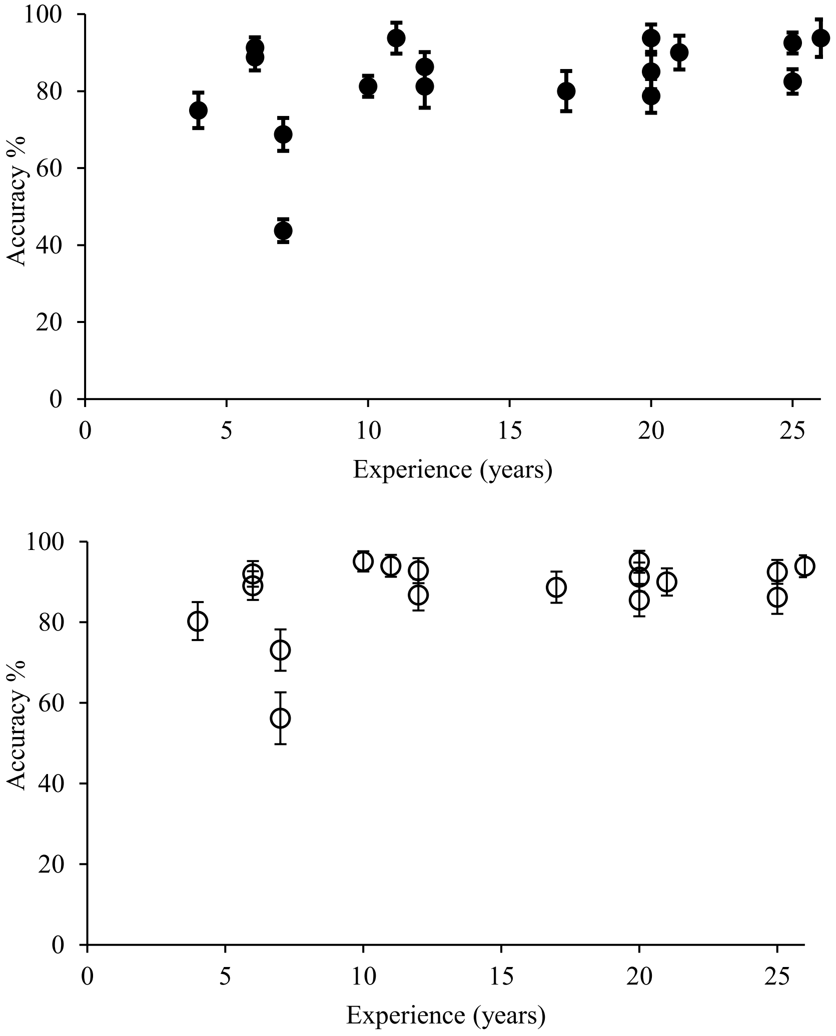


Figure 4.1. Pearson’s correlation of experience with individual percentage mean accuracy for all images (top) and for just the four study species (bottom) all images, compared with experience Error bars show ±1 standard error around the mean.

When comparing accuracy between participants grouped by self-assessed ability, no difference was found ($t_{(15)}$ = -1.32, p = 0.207) (Fig. 4.2a). In addition, a one factor ANOVA found no difference in accuracy of participants grouped according to self-assessed ability in comparison with their peers ($F_{(2,14)}$ = 0.03, p = 0.969) (Fig. 4.2b). When analysed by experience grouped by five year periods, average accuracy was highest for participants with experience of 20 years or more at 89.6%, but there were no differences between the groups ($F_{(4,12)}$ = 1.38, p = 0.297) (Fig. 4.2c). Participants who surveyed in a professional capacity (n = 14) were no more accurate than those surveying as volunteers only (*n* = 3) ($t_{(15)}$ = 0.90, p = 0.383) (Fig. 4.2d).

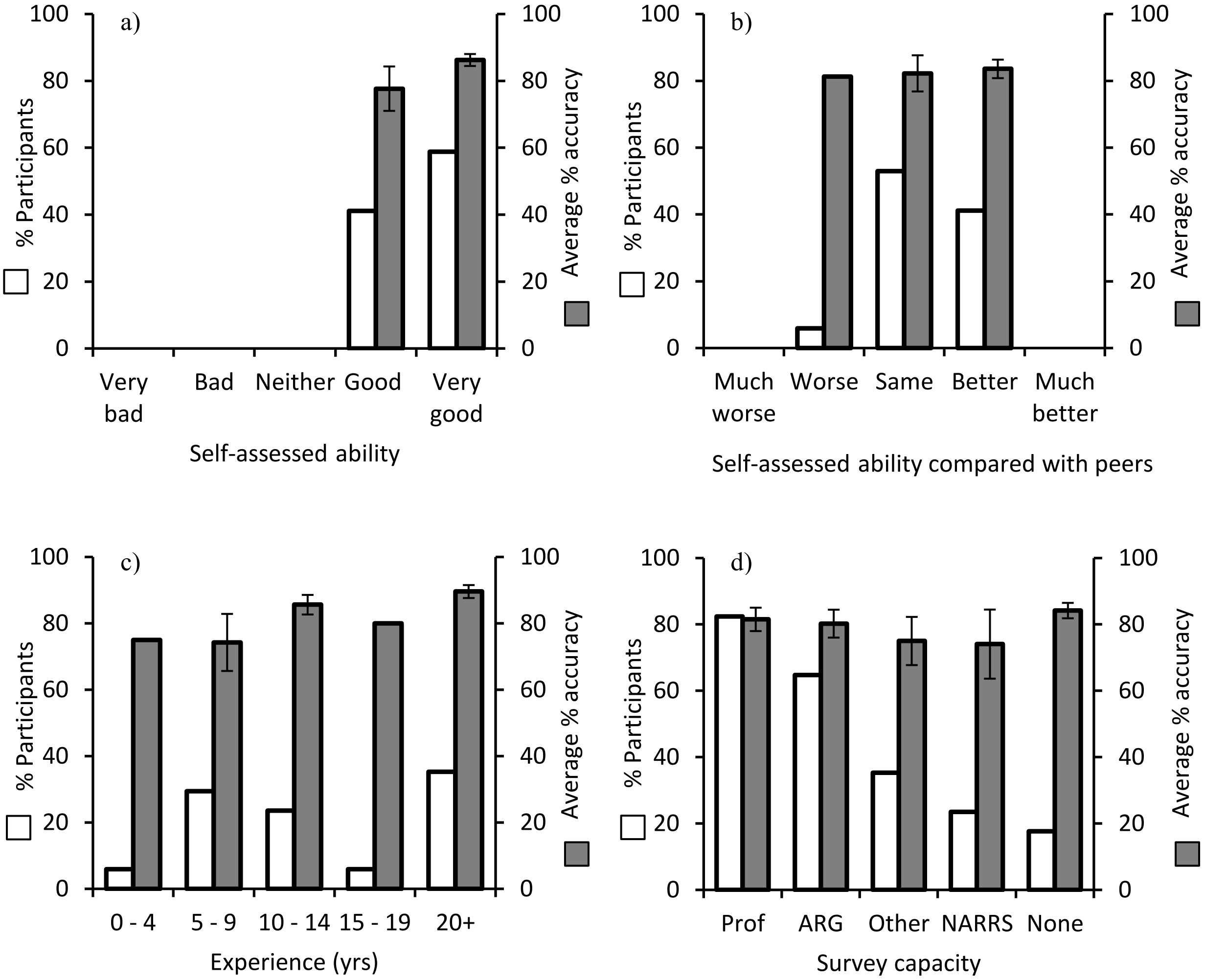


Figure 4.2. Average accuracy of participants (grey) in relation to the following factors (white): (a) self-assessed abilities; (b) self-assessed ability in comparison with peers; (c) years of experience in surveying; (d) type of surveying ("Prof" = professional, "ARG" = Amphibian and Reptile Groups, "Other" = affiliations not listed, NARRS = "National Amphibian and Reptile Recording Scheme", and "None" = no affiliation). In (d) participants total more than 100% as 12 participants surveyed in more than one capacity. Error bars show ±1 standard error around the mean.

Participant choice when grouping and naming images varied between species. When images were analysed to investigate how participants had sorted and named them, images of *T. cristatus* were grouped together most often, with relative consistency (Fig. 4.3). Conversely, grouping of *L. helveticus* images was highly variable (Fig. 4.3). A one-way ANOVA using arcsine square-root transformed scores found that the consistency with which images were grouped together varied between species ($F_{(3,76)} = 7.64$, $p < 0.001$). Tukey post-hoc test revealed that images of *T. cristatus* were grouped together more frequently than *L. helveticus* ($p < 0.001$) and *L. vulgaris* ($p = 0.007$). Moreover, participants agreed with the naming of 95.9% of *T. cristatus*, 87.6% of *I. alpestris*, 78.5% of *L. vulgaris*, and 68.8% of *L. helveticus* images. In total, 22.5% of images were named as the same species by every participant *and* in agreement with the image label. There were no images for which all participants agreed on one species name, which differed from the species named in the image.

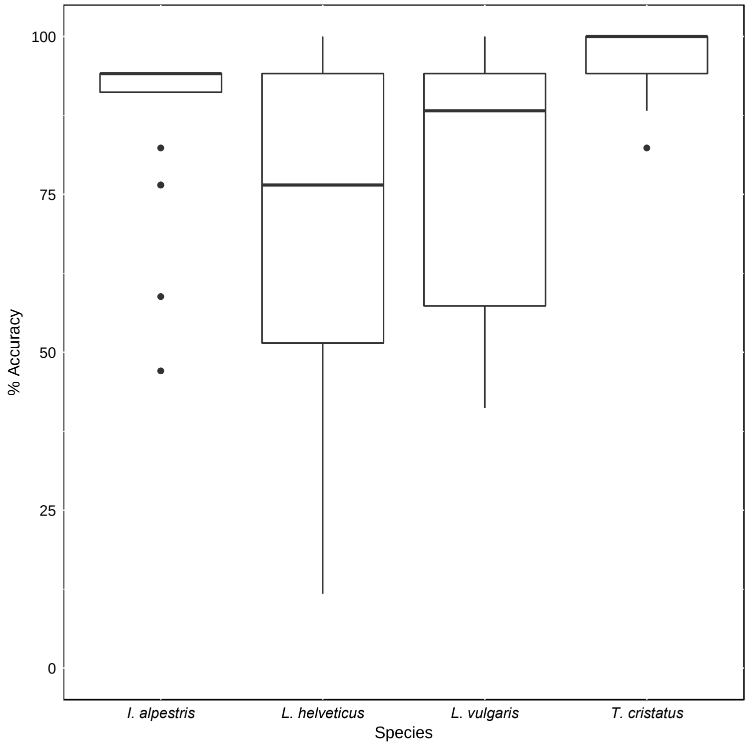


Figure 4.3. Boxplot showing accuracy rates per species. Median identification accuracy is shown by black lines in boxes, and dots are outliers.

How images were grouped and named is visualised in a confusion matrix (Fig. 4.4). The columns are the names given to the online image, and the rows are the names given by participants. Agreement on species names was highest for *T. cristatus* (96%) and differences were due to participants categorising photographs as ‘unknown’ (*n* = 8), naming images as *T. carnifex* (*n* = 4), *L. vulgaris* (*n* = 2) and *I. alpestris* (*n* = 1). Participants agreed in naming 88% of *I. alpestris* images, and grouping and agreement of names was less frequent for images of the other newts, *L. helveticus* (69%) and *L. vulgaris* (79%). Two participants created a *L. helveticus/vulgaris* hybrid group plus one participant only named ‘palmate’ *or* ‘smooth’). Furthermore, misidentification *between* these two species was notable, with approximately 1-in-8 *L. vulgaris* named as *L. helveticus*, and nearly 1-in-4 *L. helveticus* named as *L. vulgaris* (Fig. 4.4).

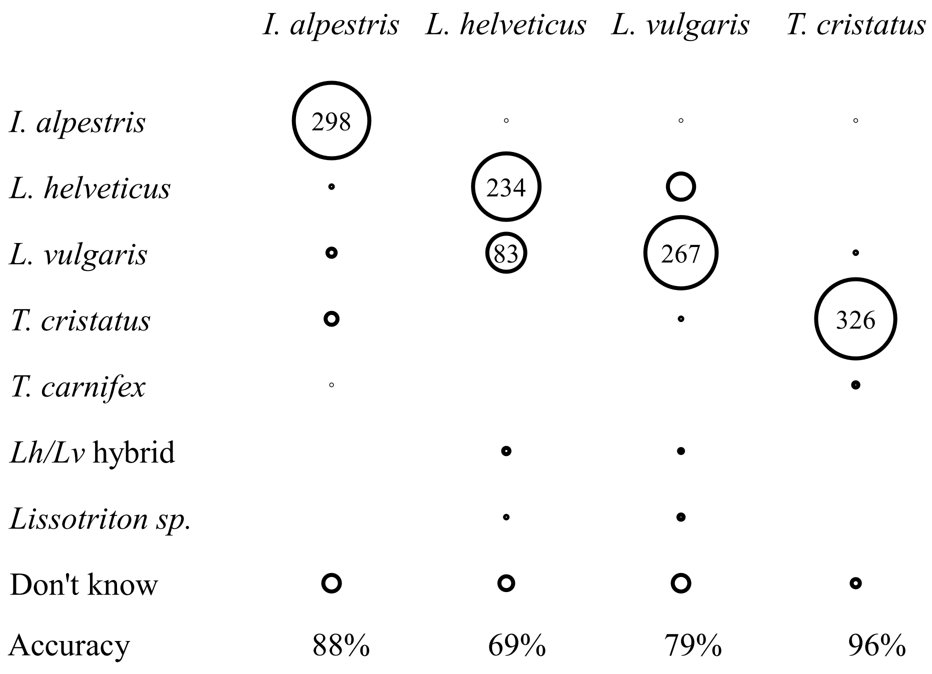


Figure 4.4. Confusion matrix comparing species in photographs, as per name assigned to the image online (columns), with species named by participants (rows). For each column (*n* = 340) percentage accuracy in naming species is shown at the bottom.

4.4.1 Characteristics visible in photographs

An effect of perspective was found (Fig. 4.5) with a one-way ANOVA ($F_{(2,77)} = 4.15$, $p = 0.019$), with a post-hoc Tukey test revealing accuracy to be significantly higher in lateral than dorsal views ($F_{(2,77)} = p = 0.031$). Average accuracy was highest for ventral views ($n = 9$) (Fig. 4.5a), but given the behaviour of newts in the wild a ventral view is uncommon unless the animal is handled. Average accuracy was highest when the whole organism was visible (Fig. 4.5b), but a one-way ANOVA found no effect of which parts of the body (namely head, head and body, or whole) were visible ($F_{(2,77)} = 0.99$, $p = 0.377$). This may seem counterintuitive, but a good example is the image that showed just the head of *T. cristatus*, yet returned 100% agreement. In comparison, the lowest score (11.8%) was for an image that also showed just a head, but of *L. helveticus*, while 88.2% participants agreed with the identification of an image showing just the head of *L. vulgaris*. However, the angle by which this *L. vulgaris* was photographed revealed spots under the 'chin', which is a diagnostic feature for this species. Accuracy was not influenced by the visibility of the defined characters of hind feet or the whole tail (Fig. 4.5c). A two-way ANOVA showed that there was no interaction between perspective and which parts of the body were visible ($p = 0.33$).

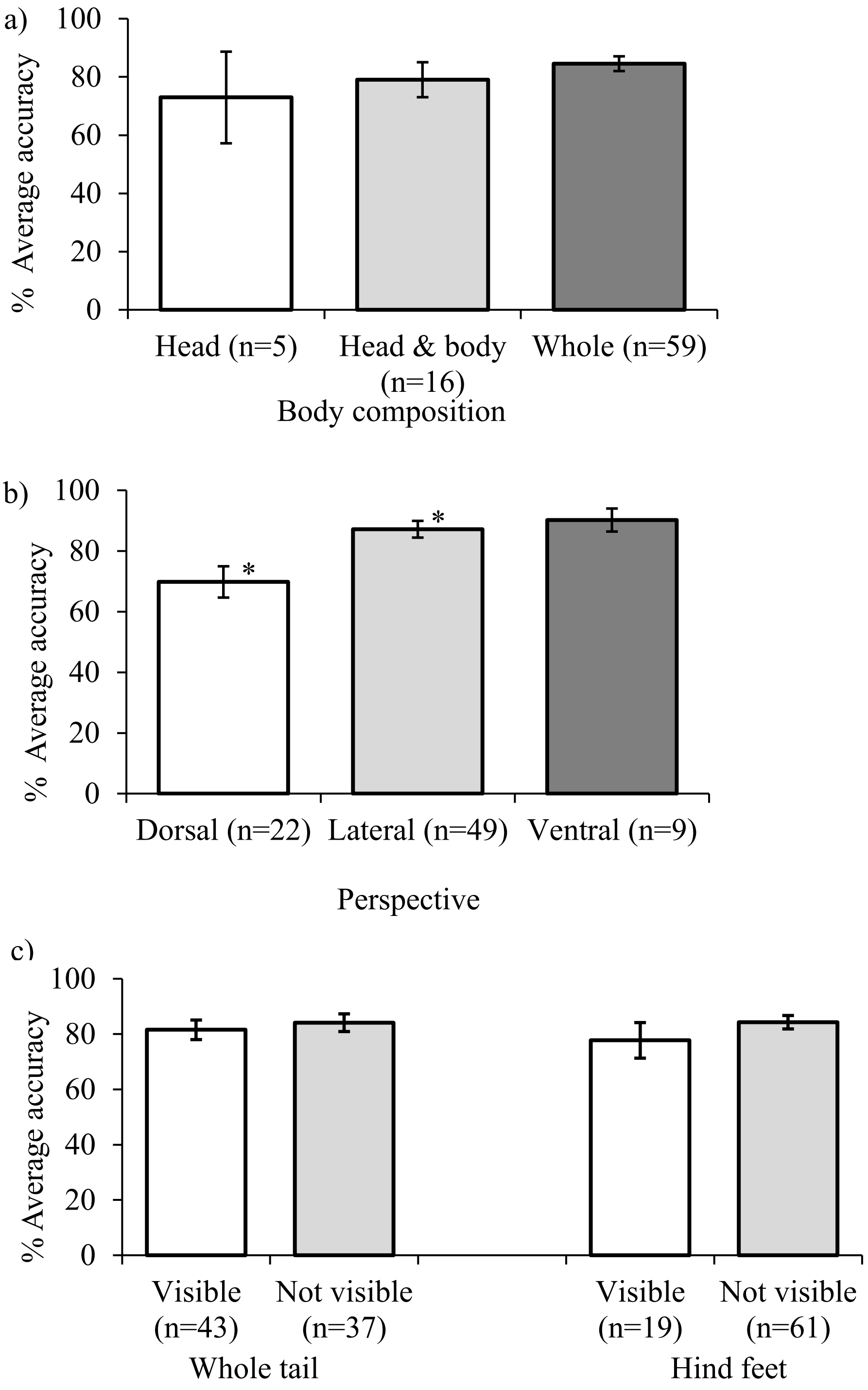


Figure 4.5. Mean accuracy relating to content and composition of images a) ventral perspective, b) whole organism, and (c) visibility of defined characteristics. Error bars show ±1 standard error around the mean, and * denotes significant difference.

In the 18 (22.5%) photographs for which all participants agreed on names, 13 of these were males exhibiting breeding characteristics. These were predominantly the ‘palm’ hind feet for *L. helveticus*, and the markings, colouration and crests for *L. vulgaris* and *T. cristatus*, all of which are diagnostic, but temporary features. The remaining five were female *T. cristatus*. Although these photographs were from different angles, the photographs included the whole body, and the distinct colourations and skin textures were visible.

4.5 Discussion

This study examined how experts grouped and named species in images downloaded from Internet sources. In this task, mean overall accuracy was 83%, and increased to 87% for when calculating accuracy in naming the four study species *only*. However, no participant was consistent in their species identification decisions, and approximately half of participants named species that were not present or made ‘don’t know’ decisions (Table 4.1). In addition, neither self-assessed abilities nor experience were indicators of individual performance. For example, although all participants regarded their identification ability as either ‘good’ or ‘very good’, individual overall accuracy ranged from 44% to 94%. This increased when only the study species were taken into account, but was still subject to broad individual differences, from 56% to 95% (Table 4.2). Participants were also limited in judging how their ability compared to their peers, and performance did not differ for those who surveyed in a professional capacity from those who surveyed solely as volunteers (Fig. 4.2).

In this task, differences between participant identification and the species named in the downloaded image were due to participants naming species not present, deciding not to assign a *species* name to a group, and confusion between the study species. In addition to the four study species, participants believed images to show *T. carnifex* (*n* = 2), hybrids of the two *Lissotriton* species (*n* = 2), and only named the genus, *Lissotriton* (*n* = 1) (Table 4.1). In addition, six participants did not assign a species name to one of their groups (Table 4.1). While approximately half of participants (*n* = 9) grouped and named just the four study species, the images within these species groups differed between participants (Table 4.1). Mean accuracy of these participants was 91%, compared with an average of 74% accuracy by those who named a group as 'don't know'. While it is possible that some participants were unable to name species, it also conceivable that they were unable to make a confident judgement from the image, and therefore refrained from doing so. When analysed by species, the consistency of identifications was also highly variable, ranging from 96% for *T. cristatus* to 69% *L. helveticus* (Fig. 4.4). All participants agreed with the name of 23% of the images used in this task, comprising the native newts *T. cristatus* (61%), *L. helveticus* (22%), and *L. vulgaris* (17%), but there was no agreement for *I. alpestris* (Table 4.2). Even when the images that were named as 'unknown or 'don't know' were removed, agreement was still low at 50%.

The variability in species identification found in this study supports previous findings that agreement between experts can be inconsistent and have limitations (Burgman et al., 2011). While additional years of experience did not improve performance (Fig. 4.1), increased expertise could result in a participant being more cautious in committing to identification (Austen et al., 2016). As surveyors, participants in this study will have substantial field experience, allowing specimens to be handled and closely observed. This method of learning has been shown to be more effective than information gleaned from books alone (Culverhouse et al. 2003), and supports the concept of ‘jizz’ (see Ellis, 2011). Although the root of the word is unknown, it is widely observed by field naturalists (Coward & Coward, 1922; Ellis, 2011) and taxonomists (Vane-Wright, 2000; Krell, 2004; Grove-White, 2007; Scharf, 2009; Williams, 2012). It corresponds with the concept of Gestalt, whereby the configuration of an object exceeds its elements and cannot be defined simply in terms of its parts (Wertheimer, 2010). In species identification, this relates to perception by which an observer can correctly name an organism without having to study its diagnostic characteristics (Ellis, 2011). For example, in this study all participants identified *T. cristatus* from an image of just a head, and this unanimous identification was probably due to the characteristic ‘warty’ skin of this species (Inns, 2009).

For novice observers, some descriptions of species characteristics can be subjective, not applicable to all populations, and sometimes only described in terms that relate to a similar, but different species (Swiderski, Zelditch, & Fink, 1998). However, experience in species identification increases the knowledge that certain features are ‘typical’ of a species, rather than being absolute, defining characteristics. For example, a recent study of mountain bongo found that accuracy was highest when comparing inly flanks, and that the inclusion of other traits (e.g. head, legs,) confounded accuracy, and those familiar with the species made fewer misidentifications (Gibbon, Bindemann, & Roberts, 2015). The importance of certain features when comparing conspecifics could aid the design of automatic identification software. Some participants in the current study noted that while certain images were adequate, they were no substitute for handling an organism. Conversely, not all observers favour handling, or are permitted to handle, specimens, and these restrictions can limit the observation of diagnostic characteristics. This study found that agreement on identification was higher in images presenting ventral views (Fig. 4.5), but this perspective is rarely experienced unless the specimen is handled.

The limitations placed on the opportunity to handle specimens may hinder accurate identification by novices. When referring to identification guides, many images are well lit with attention drawn to diagnostic features, such as secondary sexual characteristics. However, in the field newts are often observed under low light or with a torch. Furthermore, sexual dichotomy associated with breeding means that morphological differences *within* species vary throughout the year, and characteristics observable in the aquatic environment may not be visible in the terrestrial one. Of the 18 images in this task on which all participants agreed with the species name, 13 were of males in breeding condition. Such variation may confuse novice observers, but the rise of citizen science and availability of technology means that expert judgement can be sought via images. Combined judgements can have an advantage over decisions by individuals (Surowiecki, 2004; White et al., 2013; Swanson et al., 2016), but participants were asked to complete this task on an individual basis to help avoid biases that can arise in groups of interacting experts (McBride et al., 2012). Despite their experience and expertise, participants differed in their image identification, especially with *Lissotriton* species, suggesting that verification should be sought where possible, even though this may not be achievable or practical in all situations. In this study, participants agreed on the species name for less than one-in-four of all images, demonstrating the difficulty of naming an organism from a photograph alone.

This study also highlights some of the issues associated with testing identification accuracy and expert participation experimentally. By definition, expert knowledge is knowledge on a subject not commonly known (Martin et al. 2012). Consequently experts form only a small part of the general population. In addition, not all experts may be willing to undergo experimental assessments of their ability, thus limiting expert participant pools further. Participant numbers in the current study could have been increased if invites had gone beyond the targeted audience of great crested newt license holders at this meeting. However, licence holders outside of this community, such as ecological consultants, are unlikely to be sent images for identification or verification by the public. Selecting stimuli can also be problematic, as most have the potential to contain some element of error. For example, a study by Culverhouse et al. (2003) found that of specimens that had been labelled by the author and validated by an independent taxonomist, expert consensus on specimen names was just 43%. Similarly, the current study found expert agreement with species named in downloaded images to be highly variable, even though images were sourced from websites dedicated to species identification and herpetofauna (e.g. iSpot, iNaturalist, Amphibian and Reptile Groups, and Amphibian and Reptile Conservation Trust).

Regardless of the source, the extent to which the reliability of an online identification can be ascertained is linked to the perceived expertise of a person (Eiser et al., 2009). This study found that perceptions of such expertise were not linked to accuracy (Table 4.2), yet a novel observer seeking help with identification will need to decide whom to believe. These judgements may be further confounded by the way Internet search processes contain elements of crowdsourcing. This means that when an image is found following an Internet search, if that image is selected it then becomes linked to that search *term* (Pimm et al., 2015). For example, a search for *Lissotriton helveticus* also returns images of *L. vulgaris*, which is only evident from information provided after the image has been selected. Therefore, images found under specific search terms may indicate confidence in the identification. Equally, images found from the same search term could be of *different* species, depending on previous searches. It is possible that participant identification and consensus in this task may have been improved if the stimuli images had been taken from identification guides. However, analogous research on forensic face matching suggests that experiments controlled in this manner underestimate errors, as they do not account for the natural variation found in realistic settings (Megreya, Sandford, & Burton 2013).

This study was concerned with the issues of newt misidentification and the lack of demonstrable identification skills required when applying for an EPS licence. Accordingly, model species were limited to the four that licence holders could encounter when surveying in the UK. Considering that expert judgement was already variable in such a constrained context, decisions required on more speciose taxonomic groups could create greater levels of disagreement. In terms of licensing, the confusion between *L. helveticus* and *L. vulgaris* found in this study (Fig. 4.4) would not be problematic. However, this misidentification does have the potential to be misleading when assessing numbers and the distribution of those species. The effects of misidentification also go beyond individual surveys. One condition of an EPS licence is that all observations of the licensed species are submitted to a relevant recording database. In this study, participants agreed that 96% of *T. cristatus* were correctly named, and although a misidentification rate of 4% could be considered low, it could be amplified when applied to large datasets.

In summary, the results in this study suggest that consensus may be a more appropriate indicator than accuracy in species identification. Apart from a handful of well-known species, accurate identification often requires the skills of taxonomic specialists (Pimm et al., 2015). However, we found that when presented with the same stimuli, the grouping and naming of images was inconsistent across participants (Table 4.1) and species (Fig. 4.3). While there are philosophical debates around trusting experts and their decisions (Goldman, 2001; Burgman, 2015), expert judgements remain crucial for ecological monitoring. This research concurs with other studies that found that consistent identification is not linked to experience, or that expertise is the domain of the professional (Burton et al., 1999; White et al., 2014; Landrum & Mills, 2015; Austen et al., 2016). Although there are few studies directly comparing identification accuracy between experts, the available evidence reveals variable performance (e.g., Culverhouse et al., 2003). Given the heterogeneous nature of species variation, a repeat of this study with different taxon groups would discover if identification rates are comparable. Moreover, as most participants in this study surveyed in more than one capacity, a repeat with contributors that survey as professionals *or* as volunteers could build on our findings. Finally, although images from novices can provide useful information in a timely manner (Silvertown et al., 2015; Daume & Galaz, 2016), these images can present challenges to those asked to identify them. While data from large volumes of digital images can be novel, expedient and revealing, they are not necessarily as informative as the indefinable ‘jizz’ gained from field observations.

4.6 References

Arnold E.N., Burton J.A., Ovenden D. (1978). *A field guide to the reptiles and amphibians of Britain and Europe*. Collins, London.

Arnold N., Ovenden, D. (2002). *A field guide to the reptiles and amphibians of Britain and Europe*. HarperCollins, London.

Austen, G.E., Bindemann, M., Griffiths, R.A., Roberts, D.L. (2016). Species identification by experts and non-experts: comparing images from field guides. *Scientific Reports*, **6**, 33634.

Bates, A.J., Lakeman Fraser, P., Robinson, L., Tweddle, J.C., Sadler, J.P., West, S.E., Norman, S., Batson, M., Davies, L. (2015). The OPAL bugs count survey: exploring the effects of urbanisation and habitat characteristics using citizen science. *Urban Ecosystems*, **18**, 1477-97.

Blaney, R.J.P., Pocock, M., Jones, G.D. (2016). Citizen science and environmental monitoring: towards a methodology for evaluating opportunities, costs and benefits. *UK Environmental Observation Framework*.

Burgess, H.K., DeBey, L.B., Froehlich, H.E., Schmidt, N., Theobald, E.J., Ettinger, A.K., HilleRisLambers, J., Tewkesbury, J., Parrish, J.K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, **208**, 113-120.

Burgman, M.A. (2015). *Trusting Judgements: how to get the best out of experts*. Cambridge University Press, Cambridge.

Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L., Twardy, C. (2011). Expert status and performance. *PLoS One,* **6**, e22998.

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L, Maguire, L. (2011). Redefining expertise and improving ecological judgment. *Conservation Letters,* **4**, 81–87.

Burton, A.M., Wilson, S., Cowan, M. and Bruce, V. (1999). Face recognition in poor-quality video: evidence from security surveillance. *Psychological Science,* **10**, 243–248.

Chi, M.T.H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In: Ericsson, K.A., Charness, N., Hoffman, R.R. and Feltovich, P.J. eds. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, Cambridge, 167-184.

Cooper, C.B., Shirk, J., Zuckerberg, B. (2014). The invisible prevalence of citizen science in global research: migratory birds and climate change. *PLoS One,* **9**, e106508.

Coward, T.A. (1922). *Bird haunts and nature memories*. F. Warne & Co. Ltd., London.

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, **247**, 17–25.

Daume, S. and Galaz, V. (2016). “Anyone know what species this is?” - Twitter conversations as embryonic citizen science communities. *PLoS One*, **11**, e0151387.

Eiser, J., Stafford, T., Henneberry, J. and Catney, P. (2009). “Trust me, I’m a scientist (not a developer)”: perceived expertise and motives as predictors of trust in assessment of risk from contaminated land. *Risk Analysis*, **29**, 288-297.

Ellis, R. (2011). Jizz and the joy of pattern recognition: virtuosity, discipline and the agency of insight in UK naturalists’ arts of seeing. *Social Studies of Science*, **41**, 769–790.

Ericsson, K.A. (2014). Expertise. *Current Biology*, **24**, R508–R510.

Evans, R. (2008). The sociology of expertise: the distribution of social fluency. *Sociology Compass*, **2**, 281-298.

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J. Stevenson, R.D., Courtney, G.W., Van Dyk, J.K. and Ellison A.M. (2013). Next-generation field guides. *BioScience*, **63**, 891–899.

Farrington-Derby, T. and Wilson, J.R. (2006). The nature of expertise: a review.

Applied Ergonomics, **37**, 17–32.

Fowler, A., Whyatt, J.D., Davies, G. and Ellis, R. (2013). How reliable are citizen-derived scientific data? Assessing the quality of contrail observations made by the general public. *Transactions in GIS*, **17**, 488-506.

Germano, J.M., Field, K.J., Griffiths, R.A., Clulow, S., Foster, J, Harding, G. and Swaisgood, R.R. (2015). Mitigation-driven translocations: are we moving wildlife in the right direction? *Frontiers in Ecology and the Environment*, **13**, 100–105.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ*, **3**, e1303.

Goldman, A.I. (2001). Experts: which ones should you trust? *Philosophy and Phenomenological Research*, **63**, 85–110.

Grove-White, R., Waterton, C., Ellis, R., Vogel, J., Stevens, G. and Peacock, B. (2007). Amateurs as experts: harnessing new networks for biodiversity. *Lancaster University, Lancaster*, 1–179.

Hoffman, R.R. (1996). How can expertise be defined? Implications of research from cognitive psychology. In: Williams, R., Faulkner, W. and Fleck, J. eds. *Exploring Expertise*. University of Edinburgh Press, Edinburgh, 81–100.

Hoffman, R.R., Crandall, B. and Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: a case study in the methodology of cognitive task analysis. *Human Factors,* **40**, 254–276.

Inns, H. (2009). *Britain's reptiles and amphibians: a guide to the reptiles and amphibians of Great Britain, Ireland and the Channel Islands*. WILDGuides. Princeton University Press, Princeton.

Jenkins, R. White, D., Van Montfort, X. and Burton, A.M. (2011). Variability in photos of the same face. *Cognition*, **121**, 313-323.

Kapos, V., Balmford, A., Aveling, R., Bubb, P., Carey, P., Entwistle, A., Hopkins, J., Mulliken, T., Safford, R., Stattersfield, A., Walpole, M. and Manica, A. (2009). Outcomes, not implementation, predict conservation success. *Oryx*, **43**, 336-342.

Krell, F.T. (2004). Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of "morphospecies" sorting. *Biodiversity and Conservation*, **13**, 795–812.

Landrum, A.R. and Mills, C.M. (2015). Developing expectations regarding the boundaries of expertise. *Cognition*, **134**, 215–231.

Latombe, G., Pyšek, P., Jeschke, J.M., Blackburn, T.M., Bacher, S., Capinha, S.B., Costello, M.J., Fernández, M., Gregory, R.D., Hobern, D., Hui, C., Jetz, W., Kumschick, S., McGrannachan, C., Pergl, J., Roy, H.E., Scalera, R., Squires, Z.E., Wilson, J.R.U., Winter, M. Genovesi, P. and McGeoch, M.A. (2016). A vision for global monitoring of biological invasions. *Biological Conservation*, **213**, 295-308.

Lave, J., and Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge University Press, Cambridge.

Leighton, G.R.M., Hugo, P.S., 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**, 1060-1070.

Lewis, B., Griffiths, R.A. and Wilkinson, J.W. (2017). Population status of great crested newts (*Triturus cristatus*) at sites subjected to development mitigation. *Herpetological Journal*, **27**, 133–142.

Lindenmayer, D.B. and Likens, G.E. (2010). The science and application of ecological monitoring. *Biological Conservation*, **143**, 1317–1328.

Mang, T., Essl, F., Moser, D., Karrer, G., Kleinbauer, I. and Dullinger, S. (2017). Accounting for imperfect observation and estimating true species distributions in modelling biological invasions. *Ecography*, (early view online).

Martin, T., Burgman, M., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. and Mengersen, K. (2012). Eliciting expert knowledge in conservation science. *Conservation*, **26**, 29-38.

McBride, M.F., Garnett, S.T., Szabo, J.K., Burbidge, A.H., Butchart, S.H.M., Christidis, L., Dutson, G., Ford, H.A., Loyn, R.H., Watson, D.M. and Burgman, M.A. (2012). Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. *Methods in Ecology and Evolution*, **3**, 906–920.

McKinley, D.C., Miller-Rushing, A.J., Ballard, H.L., Bonney, R., Brown, H., Cook-Patton, S.. Evans, D.M., French, R.A., Parrish, J.K., Phillips, T.B., Ryan. S.F, Shanley, L.A., Shirk, J.L., Stepenuck, K.F., Weltzin, J.F., Wiggins, A., Boyle, O.D., Briggs, R.D., Chapin III, S.F., Hewitt, D.A., Preuss, P.W. and Soukup, M.A. (2016). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, **208**, 15-28.

Megreya, A.M., Sandford, A. and Burton, A.M. (2013). Matching face images taken on the same day or months apart: the limitations of photo ID. *Applied Cognitive Psychology*, **27**, 700-706.

O’Donnell, R.P. and Durso, A.M. (2014). Harnessing the power of a global network of citizen herpetologists by improving citizen science databases. *Herpetological Review*, **45**, 151–157.

Ohst, T., Gräser, Y. and Plötner J. (2013). *Batrachochytrium dendrobatidis* in Germany: distribution, prevalences, and prediction of high risk areas. *Diseases of Aquatic Organisms*, **107**, 49–59.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, **30**, 685-696.

Scharf, S.T. (2009). Identification keys, the “natural method,” and the development of plant identification manuals. *Journal of the History of Biology*, **42**, 73–117.

Silvertown, J., Harvey, M., Greenwood, R., Dodd, M., Rosewell, J., Rebelo, T., Ansine, J. and McConway, K. 2015. Crowdsourcing the identification of organisms: a case-study of iSpot. *ZooKeys*, **480**, 125–146.

Spitzen-van der Sluijs, A. (2016). Expanding distribution of lethal amphibian fungus *Batrachochytrium salamandrivorans* in Europe. *Emerging Infectious Diseases*, **22**, 1286.

Stebbins, R.A. (1977). The amateur: two sociological definitions. *The Pacific Sociological Review*, **20**, 582–606.

Stilgoe, J., Irwin, A. and Jones, K. (2006). The challenge is to embrace different forms of expertise to view them as a resource rather Opening up expert advice.

Surowiecki, J. (2004). The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, 296.

Sutherland, W.J. and Burgman, M. (2015). Policy advice: Use experts wisely. *Nature* **526**, 317–318.

Swanson, A., Kosmala, M., Lintott, C. and Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images.

Conservation Biology **30**, 520-31.

Swiderski, D., Zelditch, M. and Fink, W. (1998). Why morphometrics is not special: coding quantitative data for phylogenetic analysis. *Systematic Biology*, **47**, 508-519.

Vane-Wright, R.I. (2000). Taxonomy, methods of. *Encyclopedia of Biodiversity,* **7**, 97–111.

Wertheimer, M. (2010). A Gestalt perspective on the psychology of thinking. In: Glatzeder, B., Goel, V. and von Muller, A. eds. *Towards a Theory of Thinking*. Springer, Berlin, 49–58.

White, D., Burton, A.M., Kemp, R.I. and Jenkins, R. (2013). Crowd effects in unfamiliar face matching. *Applied Cognitive Psychology*, **27**, 769-777.

White, D., Kemp, R.I., Jenkins, R., Matheson, M. and Burton, A.M. (2014). Passport officers' errors in face matching. *PLoS One*, **9**, e103510.

Williams, R.E.A. (2012). *An Introduction to Bees in Britain*. Bees, Wasps & Ants Recording Society, 1–49.

Appendix 4.1

A list of website links to image sources (downloaded August 2014)

Ichthyosaurus alpestris

http://www.club100.net/species/M_alpestris/M_alpestris.html
http://www.thinkreptiles.com/amphibians/salamander-species-index
http://www.herpetofauna.co.uk/alpine_newt.htm
http://www.telegraph.co.uk/women/sex/valentines-day/9079601/Animal-love-Valentines-day-lovebirds-in-displays-of-affection.html?image=13
http://www.amphibian.co.uk/alpine.html
http://www.museumkiev.org/zoo/cadastre/triturus_alpestris_en.html
http://www.herp.it/indexjs.htm?SpeciesPages/TrituAlpes.htm
http://www.kentarg.org/alpine-newt-mesotriton-alpestric-formally-triturus-alpestris.html
http://www.herpetofauna.co.uk/alpine_newt.htm
http://www.freenatureimages.eu/animals/Amphibia,%20Amfibieen,%20Amphibians/Triturus%20alpestris,%20Alpine%20Newt/index.html#Triturus%20alpestris%2020%2C%20male%2C%20Alpenwatersalamander%2C%20Saxifraga-Willem%20van%20Kruijsbergen.jpg
http://www.schule-bw.de/unterricht/faecher/biologie/projekt/amphibien/ichthyosaura_alpestris.html
http://www.arc-trust.org/non-native-amphibians.html
http://www.planetepassion.eu/amphibians-in-france/Alpine-Newt-France.html
http://www.caudata.org/cc/species/Triturus/T_alpestris.shtml
http://www.iucnredlist.org/details/59472/0
http://www.caudata.org/forum/f1173-advanced-newt-salamander-topics/f24-photo-gallery-video-gallery-technique-discussion/f25-photo-video-gallery/64473-mesotriton-alpestris-apuanus.html
http://www.surrey-arg.org.uk/SARG/08000-TheAnimals/SARG2Amphibians.php
http://www.nahuby.sk/obrazok_detail.php?obrazok_id=69312
http://forum.serpenti.it/viewtopic.php?f=83&t=16224
http://www.ufz.de/index.php?en=16938

Lissotriton helveticus

http://www.caudata.org/cc/species/Triturus/T_helveticus.shtml
http://www.caudata.org/cc/species/Triturus/T_helveticus.shtml
http://inpn.mnhn.fr/espece/cd_nom/444432
http://www.surrey-arg.org.uk/SARG/08000-TheAnimals/SARG2Amphibians.php
http://www.herpetofauna.co.uk/palmate_newt.asp
http://www.devonwildlifetrust.org/species/palmate-newt
http://www.butemuseum.org.uk/natural-history-gallery/fauna/reptiles-and-amphibians/
http://www.arkive.org/palmate-newt/triturus-helveticus/image-A7572.html
http://www.arkive.org/palmate-newt/triturus-helveticus/image-A7570.html
http://www.arkive.org/palmate-newt/triturus-helveticus/image-A6802.html
http://www.arkive.org/palmate-newt/triturus-helveticus/
http://www.kentarg.org/Amphibians/palmate-newt
http://www.kentarg.org/Amphibians/palmate-newt
http://www.naturespot.org.uk/species/palmate-newt-1
http://www.devonwildlifetrust.org/species/Palmate+newt/
http://www.bbc.co.uk/nature/22120791
http://www.uk-wildlife.co.uk/palmate-newt-lissotriton-helveticus-male/
https://www.google.co.uk/search?q=lissotriton+helveticus&client=firefox-a&hs=nFT&rls=org.mozilla:en-US:official&channel=sb&source=lnms&tbm=isch&sa=X&ei=SOxsU62jDYrYPL3ZgLAL&ved=0CAYQ_AUoAQ&biw=900&bih=817#facrc=_&imgdii=_&imgrc=rPjKSXddI3EznM%253A%3BgalN_OqrOSoe1M%3Bhttp%253A%252F%252Fiucnredlist-photos.s3.amazonaws.com%252Fmedium%252F1159557300.jpg%253FAWSAccessKeyId%253DAKIAJIJQNN2N2SMHLZJA%2526Expires%253D1430413300%2526Signature%253DKhh5XwPEE2cFL%25252FPIZ3OF6UJEo%25252B4%25253D%3Bhttp%253A%252F%252Fwww.iucnredlist.org%252Fdetails%252F59475%252F0%3B480%3B320
http://www.arc-trust.org/palmate-newt.html
http://www.arc-trust.org/palmate-newt.html

Lissotriton helveticus

http://www.arkive.org/smooth-newt/triturus-vulgaris/image-A9678.html

http://carlcorbidgefieldherping.blogspot.co.uk/2011/08/evidence-of-successful-reptilian.html
http://surrey-arg.org.uk/SARG/08000-TheAnimals/SARG2Amphibians.php
http://www.euroherp.com/species/Lissotriton_vulgaris/
http://www.wildlifetrusts.org/species/smooth-newt
http://www.inaturalist.org/check_lists/15625-Karachay-Cherkess-Check-List
http://www.wildaboutbritain.co.uk/gallery/files/2/8/7/4/herplvmalehome02a.jpg
http://www.naturespot.org.uk/species/smooth-newt-0
http://www.naturalis-historia.de/cgi-bin/Seite.pl?Systematisch;Animalia;Amphibia%20-%20Lurche;3#titel
http://www.arkive.org/smooth-newt/triturus-vulgaris/image-A8789.html
http://www.arkive.org/smooth-newt/triturus-vulgaris/image-A22254.html
http://www.kentarg.org/Amphibians/smooth-newt
http://www.froglife.org/amphibians-and-reptiles/smooth-newt/
http://www.froglife.org/amphibians-and-reptiles/smooth-newt/
http://www.bto.org/volunteer-surveys/gbw/gardens-wildlife/garden-reptiles-amphibians/a-z-reptiles-amphibians/smooth-newt
http://www.froglife.org/amphibians-and-reptiles/smooth-newt/
http://www.nederlandsesoorten.nl/nsr/concept/0AHGPYIHYLLN/biology
http://froglife-frogbites.blogspot.co.uk/2012/03/froglifes-dragon-of-month-smooth-newt.html
http://www.arkive.org/smooth-newt/triturus-vulgaris/image-A10523.html
http://www.biolib.cz/en/taxonimage/id37329/

Triturus cristatus

http://www.arkive.org/great-crested-newt/triturus-cristatus/
http://commons.wikimedia.org/wiki/File:Triturus_cristatus_Sinarp_8.JPG
http://www.oxfordshire-arg.org.uk/great_crested_newt1.htm
http://www.biolib.cz/en/image/id10824/
http://www.kentarg.org/Amphibians/great-crested-newt
http://www.kentarg.org/Amphibians/great-crested-newt
http://news.bbc.co.uk/local/gloucestershire/hi/people_and_places/nature/newsid_8314000/8314899.stm
http://www.arkive.org/great-crested-newt/triturus-cristatus/image-A22844.html

http://www.theguardian.com/environment/2012/apr/01/specieswatch-newt-ponds-conservation

http://www.arkive.org/great-crested-newt/triturus-cristatus/image-A22870.html

http://www.herpetofauna.co.uk/great_crested_newt.asp

http://calphotos.berkeley.edu/cgi/img_query?enlarge=0000+0000+0610+1121

http://calphotos.berkeley.edu/cgi/img_query?enlarge=0000+0000+1210+2741

http://biodiversitatecbc-apmis.ro/new/?page=galerie

http://www.iucnredlist.org/details/22212/0

http://www.bbcwildlife.org.uk/node/3061

http://www.arkive.org/great-crested-newt/triturus-cristatus/image-A18992.html

http://www.euroherp.com/Resources/Trips/up/128-1361207113.jpg

http://speciesofuk.blogspot.co.uk/2013/07/week-21-great-crested-newt-triturus.html

http://www.biolib.cz/en/taxonimage/id10829/?taxonid=309

Appendix 4.2

Participant___________

Many thanks helping with this survey. The survey is anonymous, but we would like to gather some information about you.

Do you agree to participate in this survey? ○ Yes ○ No

1. What is your gender?

○ Female ○ Male ○ Other

2. Which age range applies to you?

○ under 18 ○ 18-24 ○ 25-34 ○ 35-44

○ 45-54 ○ 55-64 ○ 65+

3. Do you consider yourself to have normal vision?

○ Yes ○ No, but my vision is corrected to normal with glasses or contacts

○ No, I have a visual impairment

4. Do you hold a great crested newt licence?

○ Yes When did you obtain your first gcn licence?

○ No

5. How do you rate your ability to identifying amphibians found in the UK?

○ Very bad ○ Bad ○ Neither good nor bad ○ Good ○ Very good

6. Have you been involved in survey work that involves newts in the last 5 years?

○ Yes ○ No

7. If yes, is it in any of the following capacities (please tick all that apply):

○ Professional ○ With an ARG ○ NARRS

○ No affiliation Other: ☐

8. If you have surveyed in a non-professional capacity, in which area(s)?

(*E.g. Kent, Canterbury, CT1*) ____

9. Are you involved with training other people?

○ Yes ○ No

10. How do you rate your ability to identify amphibians in comparison to your peers?

○ Much worse ○ Worse ○ Same as peers ○ Better ○ Much better

Thank you very much for answering these questions.

Please now continue to the image matching task.

Shrew or kiwi? ©teenybiscuit

Chapter 5.

5 Same or different? Variation in grouping novel species with free and constrained choice.

5.1 Abstract

Categorising species involves decisions based on similarities and differences between organisms. Moreover, choices made on distinctions *between* species can be impaired by natural variation *within* species. Forensic face recognition studies show that participants sort images of the same unfamiliar face into several identities, but when informed how many identities are present, accuracy improves. Using two simple image-sorting tasks, images of four different newt species were sorted by individuals unfamiliar with these taxa. Unaware how many species were present, participants sorted images into between four and 13 species, with no individual arriving at the 'correct' solution. Participants were then informed that four species were present, and 68% of participants improved in their ability to group the same species together. Individual performance in both tasks was significantly correlated ($r = 0.81$) but varied *between* participants. In addition, *within* participant consistency in grouping different images was highly variable. These findings suggest that inter-specific variation does not override intra-specific variation when making visual same-different judgements, and that these judgements are inconsistent.

5.2 Introduction

Species observations underpin much of the monitoring, measuring and management of biodiversity (Tyre et al. 2003; Danielsen et al. 2005; Addison et al. 2015). Whether assessing change at a global scale (Jones 2011; Turak et al. 2016), predicting responses to changes in the environment (Potts et al. 2010; Dawson et al. 2011; Stephens et al. 2016) or implementing management plans (Shea et al. 2011; N'Guyen et al. 2015; McKinley et al. 2016), it is imperative that the information used is accurate. Image-based evidence is widely used in biodiversity assessment, for example when recording the presence of species (Roberts et al. 2010; Swanson et al. 2015; Boron et al. 2016), identifying individuals (Kitchen-Wheeler 2010; Horn et al. 2014; Gibbon et al. 2015; Urian et al. 2015) and evaluating inter-species dynamics (Carbone et al. 2001; Royle & Link 2006; O'Connell et al. 2010). Categorising images into usable information relies on judgements as to whether the *subjects* of the image are the same species or individual, as opposed to the images being the same or different. As well as considering the variability *of* images, such as differences in image capture devices, environmental conditions, perspective, and distance (Jenkins et al. 2011), this process also incorporates the natural variability of the species shown in the image. Organisms are not fixed in appearance: individuals vary with age, breeding condition, and health; disease or injury can cause body dysmorphia; local adaptation can result in differences between populations of the same species; and mutations can affect colouration and markings. Variability in the appearance of conspecifics has important consequences for identification. Describing and naming species provides discrete units by which living things can be studied, and involves determining patterns of similarity and difference between organisms (Vane-Wright 2000). However, the creation of discrete taxonomic units from a continuum of heritable characters (Heywood 1998;

Hey 2001), combined with the natural variation found within species, means that judgements on whether organisms are the same or different species are affected by both inter-specific and intra-specific variability.

An interesting parallel to this decision-making process is found in the study of forensic face identification, where judging stimuli as ‘same’ or ‘different’ is an experimental paradigm used widely in face recognition research (Zhang et al. 2013; Andrews et al. 2015; Sauerland et al. 2016). The study of forensic human face recognition has, until recently, focused on between-person variability and often ignored within-person variability (Bindemann & Sandford 2011; Jenkins et al. 2011). Furthermore, progress in understanding identification has been hindered by experiments using tightly controlled stimuli (Burton 2013), which have disregarded research showing that processing familiar identities is disproportionately quicker and less error-prone than processing unfamiliar ones (Johnston & Edmonds 2009; Osborne & Stevenage 2012; Burton 2013). Research using faces unfamiliar to participants found that unconstrained choice led to participants perceiving more identities than were present (Jenkins et al. 2011; Sauerland et al. 2016), while participants informed of the number of identities present were highly accurate (Andrews et al. 2015). However, exposure to greater variability in photos of the same person aided accuracy when new images of those identities were encountered later (Andrews et al. 2015).

Same-different decisions on unfamiliar subjects are common for novel observers, for example in some citizen science activities, and projects aimed at encouraging the public to participate in species identification. Observers record in different ways and refer to a range of identification resources, for example a self-organised enthusiast using a preferred guide book or website, or a novice participating in a structured citizen science project (Wal et al. 2015) where the identification material provided only includes target or common species. Unrestricted use of identification guides allows for novel observations, such as species not usually associated with a particular area or habitat, which is imperative for managing biological invasions (N'Guyen et al. 2015). Conversely, constrained choice limits the species on which identification decisions are made, but may help observers focus on particular taxa. Identification aids are designed in accordance with the desired outcomes, and here we explore differences in judgements on whether species are the same or different when choice is unconstrained and constrained.

This study investigated decisions on species being the same or different using two image-sorting tasks. Images of four newt species (family: Salamandridae) found in the UK were sorted by participants unfamiliar with these taxa, initially without knowing how many species were present (unconstrained), and then repeated knowing that four species were present (constrained). This builds on the study in Chapter 3, where same-different judgements were made on images presented as simultaneous pairs under optimised conditions, and were manipulated to create a range of similarities and differences. In this study, the stimuli are images downloaded from the Internet and incorporate natural variability (known as ambient images), therefore adding ecological validity to the study.

5.3 Method

This research was approved by the Ethics Committee of the School of Anthropology and Conservation at the University of Kent, and conducted in accordance with the ethical guidelines of the British Psychological Society. Participants were recruited via the School of Anthropology and Conservation at the University of Kent (12 female, 10 male, range = 25-70 years). All participants gave informed consent and reported good vision or corrected-to-normal.

5.3.1 Procedure

A set of 80 images, comprising four newt species found in the UK, was used as stimuli. These were the non-native alpine newt (*Ichthyosaura alpestris*), the great crested newt (*Triturus cristatus*), and two smaller newts of the same genus, the palmate newt (*Lissotriton helveticus*) and the smooth newt (*L. vulgaris*). Images for each species were found via Google© search engine by using the Latin binominal in the ‘images’ option. Twenty unique images were selected for each species, printed and laminated. These images included both males and females, although few of the labels accompanying the images noted the sex. For the task, this list of images was randomised and numbered, and these randomised numbers were assigned to the 80 images to avoid participants detecting a number pattern on the reverse of the stimuli.

5.3.2 Participants

Volunteers were recruited via the University of Kent, through social media, and by word of mouth, with participation being restricted to those unfamiliar with newts. Participants (*n* = 22) were given the set of 80 target images and asked to sort them into groups of the same species, irrespective of gender, and informed that any images containing more than one newt showed the same species. These were the only instructions given and participants were free to create as many or as few groups as they wished, with no time restrictions. On completion, the total number of groups was noted, then these groups were numbered arbitrarily and the relevant group number assigned to the corresponding image. Participants were then informed that the 80 images comprised four different species, and asked to repeat the task but constraining choice to four groups, with no limit to the number of images per group, and no time restrictions. These four groups were numbered arbitrarily, and again each group number (one to four) was assigned to the corresponding images for each participant. Images were shuffled between each task, and participants were supervised.

5.3.3 Analyses

For unconstrained choice, the number of different species recognised (the number of groups created) by each participant was counted, and a t-test comparing the number of perceived species with the number of species present, was performed using arcsine square-root data. For constrained choice, groups were analysed for species composition, and the associated label was assumed to be the correct name for the species in that image (see Chapter 4). As participation was sought from people unfamiliar with the target species, participants were not asked to name the newt species in their four chosen groups as in Chapter 4. Instead, each group was named after the majority species and this was considered the ‘correct’ species. Images of non-majority species in each group were counted as ‘incorrect’. Groups were analysed for both the number of different species present, and the percentage composition of those species. To compare performance in grouping the same images under both unconstrained and constrained conditions we used a similarity metric, the normalized mutual information (NMI_{max}) (Vinh et al. 2010). For convenience, the subscript max is omitted and NMI used hereafter. The NMI compares two clusterings, giving a numerical value between zero and one, with larger values indicating better agreement.

Similarities between the true species and the groupings obtained under unconstrained and constrained tasks were compared using a paired *t*-test. Calculations were done using R (version 3.4.0). Constrained choice was visualised in a fluctuation diagram (Pilhöfer et al. 2012) produced using the R package extracat.

5.4 Results

5.4.1 Unconstrained choice

When no restriction was placed on how many groups images could be sorted into, participants created between four and 13 groups (mean 6.7; median 6; mode 5) (Fig. 5.1). No participant perceived fewer, but most perceived more, than the four species present (Fig. 5.1). Three participants (13.6%) sorted the images into just four groups, but none of these groups comprised all 20 images of a single species. A one-sample t-test revealed that the number of perceived species was higher than the four species present ($t_{(22)} = 12.76$, $p < 0.001$).

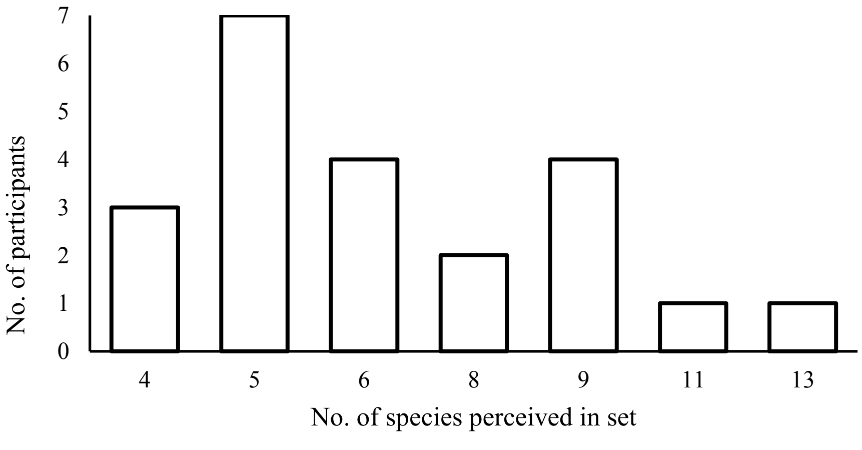


Figure 5.1. Summary of participant perception of the number of different species represented in the stimuli. Unconstrained choice when sorting images resulted in the majority of participants perceiving more species than the four present.

5.4.2 Constrained choice

Participants were then informed that the stimuli contained four species and their choice was constrained to sorting the 80 images into four groups only. The number of images in each group ranged from 9 − 39 (SD ± 5.73), no participant sorted their images into four groups of 20, and although 9 groups did contain 20 images, only one of these groups comprised *all* images of *one* species, namely *T. cristatus* (participant 22). As well as variability in the number of images in each group, the combination of the species that participants selected for each group was variable *within* and *between* participants. For six participants, their group choices resulted in at least one group *not* having a majority species, meaning that maximum percentage composition of images within one group was shared by two different species. The remaining 16 participants had a majority species in each of their four groups, nine of whom had four *different* majority species, and seven with two of their four groups having the *same* majority species (*L. helveticus* (n = 4) and *T. cristatus* (n = 3)). These sixteen participants varied in how they combined different species in each group, with four graphs showing which ‘incorrect’ species were associated with the ‘correct’ (majority) species (Fig. 5.2). For example, when *I. alpestris* was the majority species in the group (n = 16), *L. helveticus* was not present at all, with *L. vulgaris* in 11 of those groups, and *T. cristatus* in six of them (Fig. 5.2). Conversely, when *L. helveticus* was the majority species (n = 20), *I. alpestris* was included as an incorrect species in 14 of those groups, *T. cristatus* in five groups, but *L. vulgaris* appeared in *every* group (Fig. 5.2).

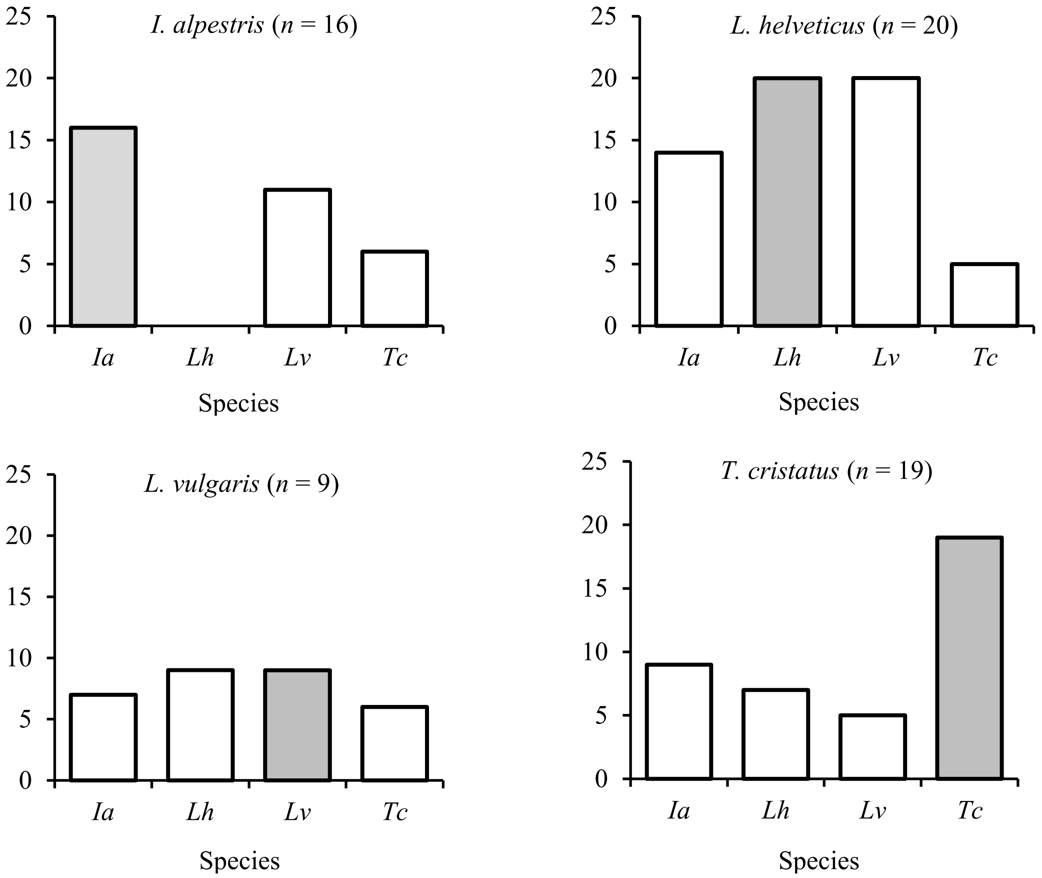


Figure 5.2. Association between majority species and other species selected to form part of the same group. When the title species was the majority in the group (grey), these graphs show on how many occasions other species were included in the same group. For example, when *I. alpestris* was the majority species, *L. vulgaris* was included in 11 groups and *T. cristatus* in six groups, but not *L. helveticus*

When analysed by the number of *images*, as opposed to the number of *species* (Fig. 5.2), the majorities of *I. alpestris* and *T. cristatus* in their respective groups are more apparent (Fig. 5.3). When *I. alpestris* was the majority species, it averaged 87.3% of the group, with the ‘incorrect’ species of *L. vulgaris* and *T. cristatus* accounting for 8.1% and 4.6% respectively. *T. cristatus* as majority species averaged 82.2%, with *I. alpestris* (5.5%), *L. helveticus* (4.3%), and *L. vulgaris* (8.0%). However, *Lissotriton* species averaged lower majorities, with the other *Lissotriton* forming a large part of the ‘incorrect’ species. *L. helveticus* averaged 54.9% as the majority species, with *L. vulgaris* averaging 36.9% of the incorrect species, plus *I. alpestris* (5.1%) and *T. cristatus* (3.1%). Similarly, *L. vulgaris* averaged 53.5% as the majority species, with *L. helveticus* averaging 30.8% of the incorrect species, and *I. alpestris* (7.3%) and *T. cristatus* (8.4%).

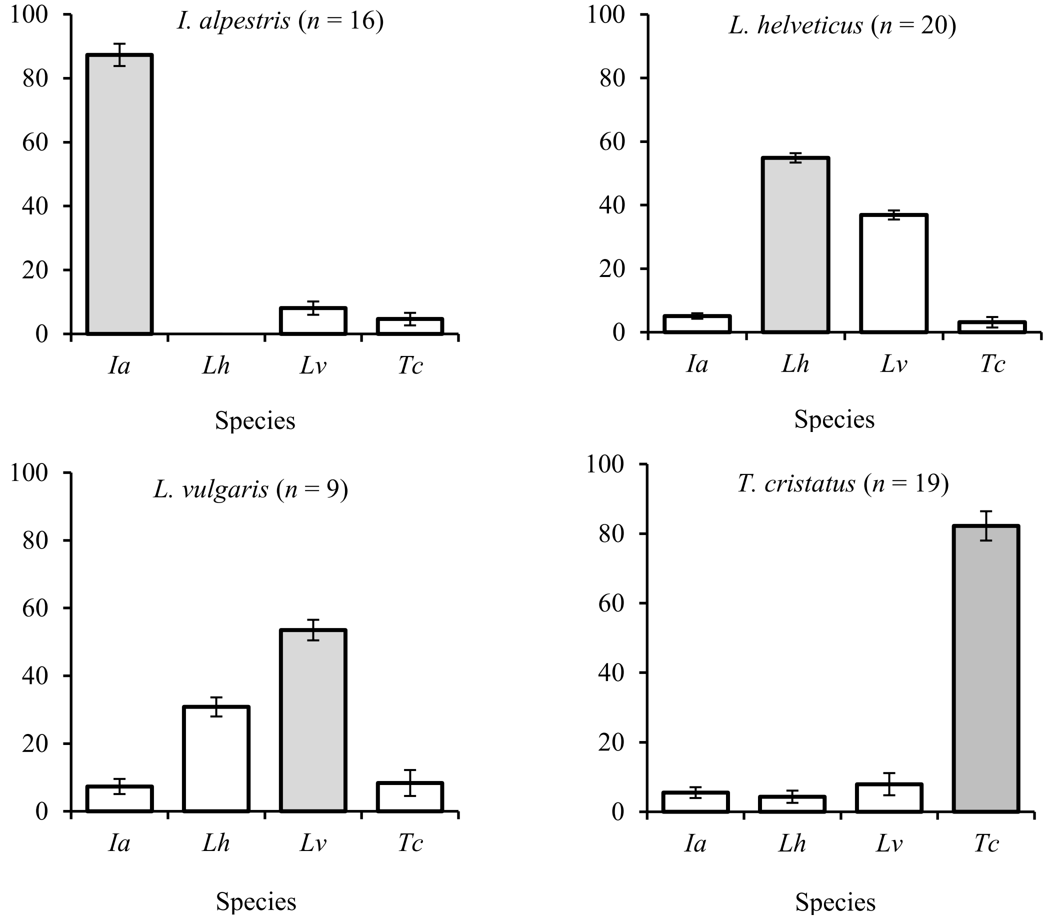


Figure 5.3. Mean percentage composition of groups per majority species. When the title species was the majority in the group, these graphs show the percentage of species included in those groups (title species in grey). For *I. alpestris* and *T. cristatus* there were a small percentage of other species, but *L. helveticus* and *L. vulgaris* show a higher reciprocal inclusion. Error bars represent ±1 standard error.

A fluctuation diagram (Fig. 5.4) visualises the variability in choice when participants sorted images into four groups. Rows represent how the 20 images of each species were grouped, and columns represent which species were present in the four groups that participants created under constrained conditions. Larger rectangles indicate a level of accuracy in recognising images as the same species. For example, participant 3 has large rectangles for *Ia* and *Tc*, which are the only ones in the column, showing that this participant chose two groups that contained just one species, namely *I. alpestris* and *T. cristatus*. The rows show that images of *I. alpestris* and *T. cristatus* were present in other groups ($n = 2$ and $n = 1$ respectively), but in small numbers as represented by small rectangles. In comparison, participant 4 has rectangles of various sizes in all boxes, showing that they have included images of all four species within each of their four groups.

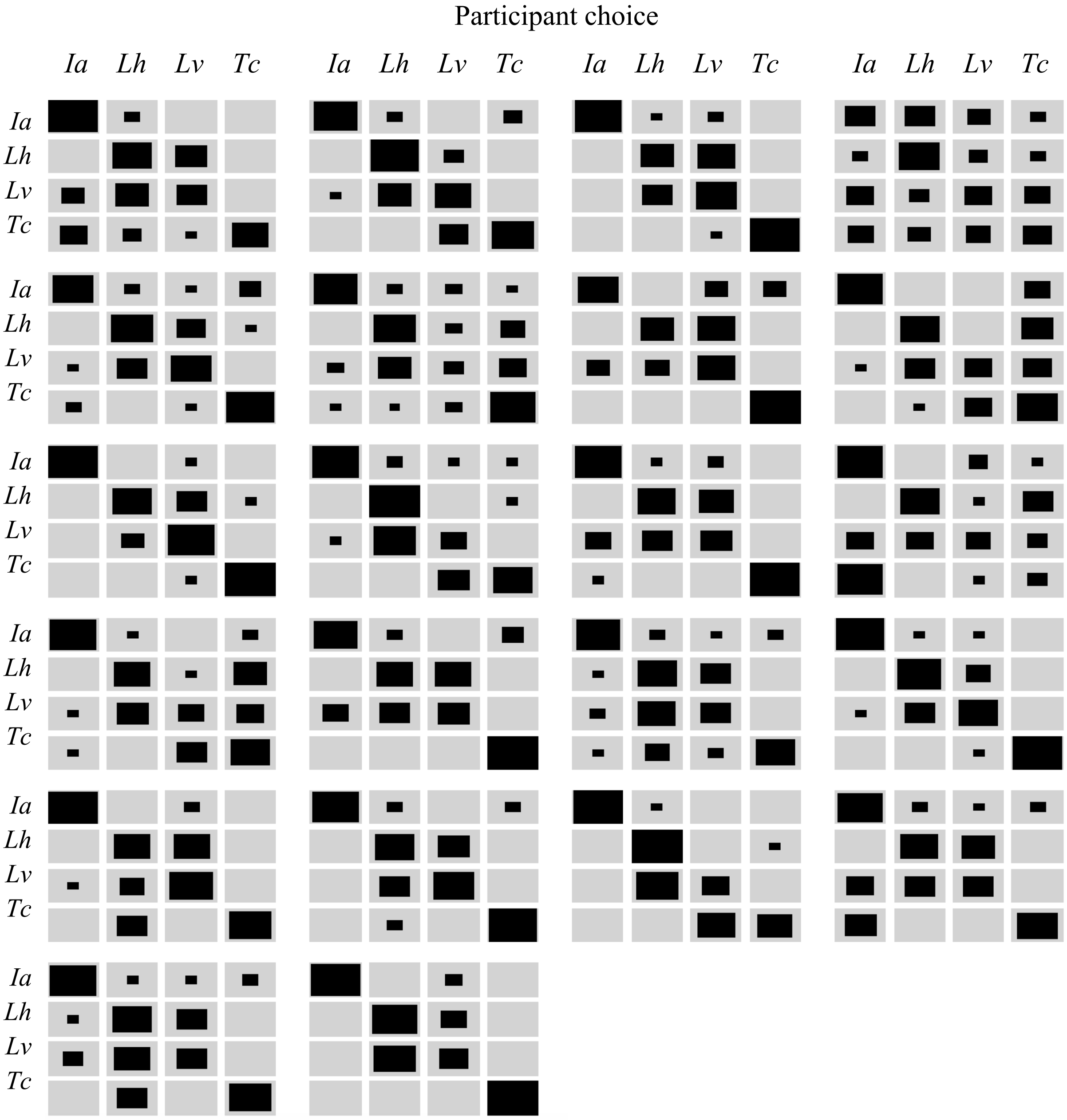


Figure 5.4. A fluctuation diagram showing species as named in images (rows) with participant choice of species (columns) for each participant (1 to 22 reading left to right along rows) for constrained choice. The larger the black rectangle within the grey box, the more accurate the participant has been in grouping the same species together.

Table 5.1 shows the similarity between the true species and the participants' allocation of species to groups under unconstrained choice (NMI_U) and constrained choice (NMI_C). Fifteen participants (68%) improved their performance under constrained choice, with Participant 3 making a notable improvement (Table 5.1), and there was a significant increase in similarity under constrained choice (mean increase = 0.058, s.e. = 0.0196, $p < 0.01$). The greatest similarity was for participant 16 (NMI_{UC}= 0.845) who chose four groups for the unconstrained element of the task, and only made five changes for constrained groups. Conversely, Participant 4 made many changes to group choices (NMI_{UC}= 0.106), as well as performing badly in the first part of the task with (NMI_{UC}= 0.152) and worse in the second (NMI_{UC}= 0.074). A Pearson product-moment correlation revealed that NMI values for both parts of the task were correlated ($r = 0.81$, $p < 0.001$), showing that individual performance was similar in both parts of the task.

Table 5.1. Measure of similarity for unconstrained choice with species (NMI_U), constrained choice with species (NMI_C), change in performance, and comparison of grouping images in both tasks (NMI_{UC}). Increase in value from unconstrained to constrained indicates improved grouping of the same species, and thus change in performance is positive (+).

Participant	Unconstrained with species	Constrained with species	Change in performance	Unconstrained with constrained
1	0.226	0.373	+	0.373
2	0.510	0.500	−	0.679
3	0.361	0.615	+	0.473
4	0.152	0.074	−	0.106
5	0.272	0.433	+	0.300
6	0.299	0.293	−	0.252
7	0.476	0.501	+	0.475
8	0.389	0.363	−	0.642
9	0.657	0.680	+	0.679
10	0.426	0.519	+	0.368
11	0.315	0.549	+	0.457
12	0.143	0.291	+	0.115
13	0.377	0.354	−	0.492
14	0.506	0.531	+	0.576
15	0.358	0.318	−	0.613
16	0.608	0.640	+	0.845
17	0.462	0.544	+	0.581
18	0.492	0.582	+	0.765
19	0.537	0.628	+	0.606
20	0.471	0.381	−	0.587
21	0.324	0.395	+	0.462
22	0.619	0.700	+	0.710

Analysis by species showed no uniformity in how participants spread images of species across their four different groups (Fig. 5.5). For example, *T. cristatus* was the only species that participants ($n = 3$) grouped all 20 images together, but two of these participants included additional images, and only Participant 22 had solely *T. cristatus* images in one group (see Fig. 5.4). This resulted in 3.4% of images of the same species being grouped together in one group, 42.1% spread across two groups, 35.2% across three groups and 19.3% across four groups. On average, *I. alpestris* and *L. vulgaris* were spread across 3.0 groups (median = 3, mode = 3), and *L. helveticus* and *T. cristatus* were spread across 2.4 groups (median = 2, mode = 2) (Fig. 5.5).

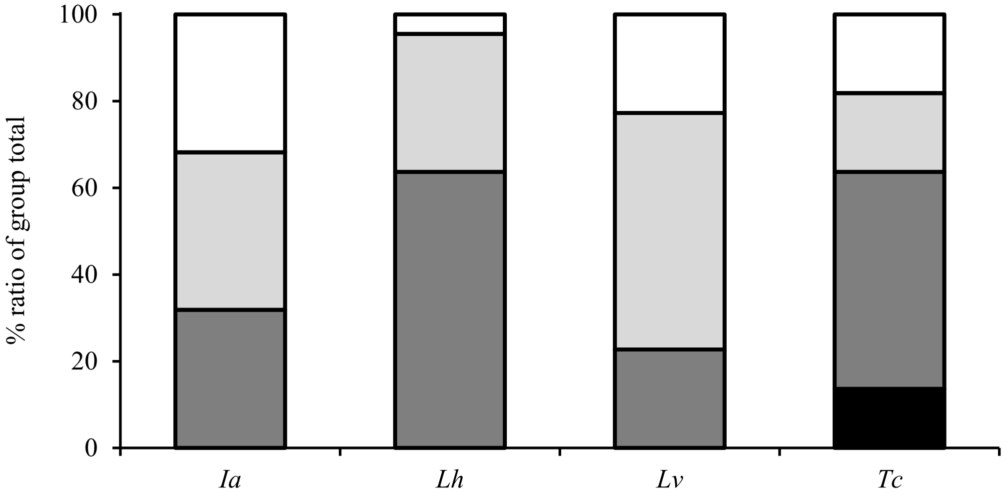


Figure 5.5. Number of groups that participants (*n* = 22) split one species across. Each column shows the percentage of participants that allocated that species to one (black), two (dark grey), three (light grey) or four (white) different groups.

5.5 Discussion

This study examined the ability of novice participants (*n* = 22) to differentiate between four species with a simple image-sorting task. Using 80 uncontrolled images of newt species, participant perception of the number of species was higher than the four present, and ranged from four to 13 (Fig. 5.1). When the task was repeated with the knowledge that four species were present, 68% of participants improved their performance (Table 5.1), and grouping of images remained variable both within and between participants. Individual performance was correlated across both tasks suggesting that categorising species on visual cues is linked to individual ability, although constrained choice resulted in improved species grouping overall. This study found variability between participants (Fig. 5.4), that all species were mistaken for another at some point (Fig. 5.2), but not necessarily in equal measure (Fig. 5.3), and that the how species were distributed across different groups was not uniform (Fig. 5.5).

Categorising species involves deciding whether organisms are the same or different, but distinction *between* species can be impaired by natural variation *within* species. The grouping of different species as the same species was most prevalent between the two newts in the genus *Lissotriton*, which may be expected in images of the most closely related newts, but confusion was observed across all species in a non-uniform way (Figs. 5.3, 5.4 and 5.5). This may have also been compounded by these taxa exhibiting sexual dichotomy. Secondary sexual characteristics are most prominent in males during the breeding season. However, these characteristics are not visible from all angles, or necessarily evident when the newt is out of water. Even if novice observers were aware of these factors, images are taken at different times of year and in a variety of contexts, and the even same individual can look different across the year, especially males. The presence of *both* inter- and intra-species variability may have made the images too difficult to interpret, but natural variability (see Jenkins et al. 2011), plus differences between image captures devices, is representative of the limitations encountered in field studies (Francesco et al. 2013; Swanson et al. 2015). Forensic face recognition studies show that participants sort images of the same unfamiliar face into several identities, but when informed how many identities are present, accuracy improves (e.g. Andrews et al. 2015). Notably, only one participant

(Participant 8 in Appendix 5.1) followed the pattern seen in face recognition studies, whereby images sorted into numerous groups in unconstrained choice are collated into fewer groups *without* those smaller groups changing. In this study, the nine groups created by Participant 8 were condensed into four groups with none of those groups changing, while all other participants split at least one of their unconstrained groups over more than one group in their constrained choice (Appendix 5.1) .Initially, research into unfamiliar face matching presented the viewer with 10 possible matches to the target identity, but more recent studies have used simple pairs of faces (Burton, White & McNeill). This has been as simultaneously presented pairs (as seen in chapters 2 and 3), or by using just two identities (Andrews et al.2015; Jenkins et al. 2011; Zhou & Mondloch 2016). A repeat of the current study starting with two species, then increasing, would explore whether the *number* of species used as stimuli affects performance. In addition, repeating the constrained sorting task would investigate within participant consistency (Bindemann, Avetisyan & Rakow 2012), and whether exposure to variable photos of the same species improved subsequent choices (Andrews et al. 2015).

Participants were unfamiliar with the model species, and just as unfamiliar faces are not processed as faces (Megreya & Burton 2006), it could follow that the stimuli were processed as animate objects rather than species. Differentiation between objects is an everyday behaviour (Lampert & Nickisch 2009), and knowing where same-different judgements become problematic would be of interest. Apart from one participant, all changed the composition of their original (unconstrained) groups, with some making few changes, while others split images initially perceived as the ‘same’ across four different groups under constrained choice (Appendix 5.1). Asking participants to explain what made the species same or different could highlight how images were being categorised, and which visual elements were important in making decisions (e.g. shape, colour, and markings). This study could also be repeated with images that have been manipulated to show subjects that correspond in aspect, scale and context, to investigate if standardising ambient images improved performance.

This task considered stimuli images to be correctly identified. However, they may incorporate an element of error, and the difficulties of testing species identification experimentally are discussed in Chapter 4. However, the task in Chapter 4 used the same stimuli as this task, and those expert participants did not reach consensus on the majority of the images, meaning that even those familiar with these taxa cannot agree on species names from images alone. Experts have proven skills or are perceived as more knowledgeable, and are therefore approached for advice (Chi 2006). While some experts are aware of their limitations (Chi 2006), and can be cautious in committing to identification from images (Chapter 2), there is some expectation that experts should be able, or willing, to accurately identify or validate species without a specimen. Chapter 2 showed that even under optimised conditions, there is uncertainty in accurate species identification, and although the images used in this task contain more ‘noise’, they are more representative of the vast amount of images being gathered by electronic means and used in citizen science projects (e.g. Swanson et al. 2015). However, an initial sorting of images may provide an important ‘first step’, similar to parataxonomy, as the increase in image related technology used to conserve biodiversity (Pimm et al. 2015), is asking more of expert communities that are limited by size and time constraints.

Although unconstrained choice creates opportunity for novel observations, intra-species variation leads to errors in recognising how many species are present. Even the constrained conditions of choice between just four species from the same family resulted in confusion between species. Furthermore, the species that comprised the ‘non-majority’ part of the four chosen groups were not reciprocal *between* species and choice was highly variable between participants. This simple sorting task shows that ability to group images by species was consistent for both tasks, but constrained choice showed overall improvement in performance. This suggests that image sorting by participants unfamiliar with the subject taxa be improved by limiting choice, but needs further exploration. In practice, the task of identification lies somewhere between constrained and unconstrained. Species that are new to science are still being discovered, both in the field and during similar sorting tasks in collections, such as in herbaria. Furthermore, factors such as climate change affect species distribution, resulting in species arriving in areas where they were previously unknown, plus the detection of novel, invasive species.

This task shows that same-different judgements are highly variable, both between individual participants and across species within the same family. Future work on different taxa and repetition of the constrained part of the task could provide further insight to patterns in same-different judgements and consistency in those choices.

5.6 References

Addison, P.F.E., Flander, L.B. and Cook, C.N., (2015). Are we missing the boat ? Current uses of long-term biological monitoring data in the evaluation and management of marine protected areas. *Journal of Environmental Management*. **149**, 148–156.

Andrews, S., Jenkins, R., Cursiter, H. and Burton, A.M. (2015). Telling faces together: learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, **68**, 2041-2050.

Bindemann, M., Avetisyan, M. and Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, **18**, 277.

Bindemann, M. and Sandford, A., (2011). Me, myself, and I: different recognition rates for three photo-IDs of the same person. *Perception*, **40**, 625-627.

Boron, V. Tzanopoulos, J., Gallo, J., Barragan, J., Jaimes-Rodriguez, L., Schaller, G. and Payán, E., (2016). Jaguar densities across human-dominated landscapes in Colombia: the contribution of unprotected areas to long term conservation. *PLoS One*, **11**, e0153973.

Burton, A.M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, **66**, 1467-1485.

Burton, A.M., White, D. and McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods,* **42**, 286-291.

Carbone, C., Christie, S., Conforti, K., Coulson, T., Franklin, N., Ginsberg, J.R., Griffiths, M., Holden, J., Kawanishi, K., Kinnaird, M. and Laidlaw, R. (2001). The use of photographic rates to estimate densities of tigers and other cryptic mammals. *Animal Conservation*, **4**, 75–79.

Chi, M.T.H (2006). Two approaches to the study of experts' characteristics. In: Ericsson, K.A., Charness, N., Hoffman, R.R. and Feltovich, P.J. eds. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, Cambridge, 21-30.

Danielsen, F., Burgess, N.D. and Balmford, A. (2005). Monitoring matters: examining the potential of locally-based approaches. *Biodiversity and Conservation*, **14**, 2507–2542.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. and Mace, G.M. (2011). Beyond predictions: biodiversity conservation in a changing climate. *Science,* **332**, 53–58.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ,* **3**, e1303.

Horn, R.C. Van Zug, B., LaCombe, C., Velez-Liendo, X. and Paisley, S. (2014).

Human visual identification of individual Andean bears *Tremarctos ornatus*. *Wildlife Biology*, **20**, 291–299.

Jenkins, R. White, D., Van Montfort, X. and Burton, A.M. (2011). Variability in photos of the same face. *Cognition*, **121**, 313-323.

Johnston, R.A. and Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: a review. *Memory*, **17**, 577-596.

Jones, J.P.G. (2011). Monitoring species abundance and distribution at the landscape scale. *Journal of Applied Ecology*, **48**, 9–13.

Kitchen-Wheeler, A-M. (2010). Visual identification of individual manta ray (*Manta alfredi*) in the Maldives Islands, Western Indian Ocean. *Marine Biology Research*, **6**, 351–363.

Lampert, C., Nickisch, H. and Harmeking, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In: *Computer Vision and Pattern Recognition*, IEEE Conference, 951-958.

Li, M., Chen, X., Li, X., Ma, B. and Vitanyi, P.B.M. (2004). The Similarity Metric. IEEE Transactions on Information Theory, **50**, 3250–3264.

McKinley, D.C., Miller-Rushing, A.J., Ballard, H.L., Bonney, R., Brown, H., Cook-Patton, S., Evans, D.M., French, R.A., Parrish, J.K., Phillips, T.B., Ryan. S.F, Shanley, L.A., Shirk, J.L., Stepenuck, K.F., Weltzin, J.F., Wiggins, A., Boyle, O.D., Briggs, R.D., Chapin III, S.F., Hewitt, D.A., Preuss, P.W. and Soukup, M.A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, **208**, 15-28.

Megreya, A.M. and Burton, A.M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. Memory and Cognition, **34**, 865–876.

N’Guyen, A., Hirsch, P.E., Adrian-Kalchhauser, I. and Burkhardt-Holm, P. (2016). Improving invasive species management by integrating priorities and contributions of scientists and decision makers. *Ambio*, **45**, 280.

O'Connell, A.F., Nichols, J.D. and Karanth, K.U. eds. (2010). *Camera traps in animal ecology: methods and analyses*. Springer Science and Business Media.

Osborne, C.D. and Stevenage, S.V. (2013). Familiarity and face processing, *The Quarterly Journal of Experimental Psychology*, **66**, 108-120.

Pilhöfer, A., Gribov, A. and Unwin, A. (2012). Comparing clusterings using Bertin’s idea. *IEEE Transactions on Visualization and Computer Graphics,* **18**, 2506-2515.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, **30**, 685-696.

Potts, S.G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution*, **25**, 345–353.

Roberts, D.L., Elphick, C.S. and Reed, J.M. (2010). Identifying anomalous reports of putatively extinct species and why it matters. *Conservation biology,* **24**, 189–196.

Rovero, F., Zimmerman, F., Berzi, D. and Meek, P. (2013). “Which camera trap type and how many do I need?” A review of camera features and study designs for a range of wildlife research applications. *Hystrix, the Italian Journal of Mammalogy,* **24**, 148-156.

Royle, J., and Link, W. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835-841.

Sauerland, M., Sagana, A., Siegmann, K., Heiligers, D., Merckelbach, H. and Jenkins, R. (2016). These two are different. Yes, they’re the same: choice blindness for facial identity. *Consciousness and Cognition*, **40**, 93–104.

Shea, C.P., Peterson, J. T., Wisniewski, J. M. and Johnson, N. A. (2011). Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.

Stephens, P.A, Mason, L.R., Green, R.E., Gregory, R.D., Sauer, J.R., Alison, J., Aunins, A., Brotons, L., Butchart, S.H., Campedelli, T. and Chodkiewicz, T. (2016). Consistent response of bird populations to climate change on two continents. *Science*, **352**, 84-87.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. and Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, **2**, 150026.

Turak, E., Turak, E., Harrison, I., Dudgeon, D., Abell, R., Bush, A., Darwall, W., Finlayson, C.M., Ferrier, S., Freyhof, J., Hermoso, V. and Juffe-Bignoli, D. (2016). Essential Biodiversity Variables for measuring change in global freshwater biodiversity. *Biological Conservation*, **213**, 272-279.

Tyre, A., Tenhumberg, B. and Field, S. (2003). Improving precision and reducing bias in biological surveys: estimating false- negative error rates. *Ecological Applications*, **13**, 1790-1801.

Urian, K., Gorgone, A., Read, A., Balmer, B., Wells, R.S., Berggren, P., Durban, J., Eguchi, T., Rayment, W. and Hammond, P.S. (2015). Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Marine Mammal Science*, **31**, 298–321.

Van der Wal, R., Anderson, H., Robinson, A., Sharma, N., Mellish, C., Roberts, S., Darvill, B. and Siddharthan, A. (2015). Mapping species distributions: A comparison of skilled naturalist and lay citizen science recording. *Ambio*, **44**, 584.

Vane-Wright, R.I. (2000). Taxonomy, methods of. *Encyclopedia of Biodiversity*, **7**, 97–111.

Vinh, N.X., Epps, J. and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, **11**, 2837-2854.

Vitányi, P.M., Balbach, F.J., Cilibrasi, R.L. and Li, M. (2009). Normalized information distance. In: *Information theory and statistical learning*. Springer US, 45-82.

Zhang, R., Hu, Z., Debi, R., Zhang, L., Li, H. and Liu, Q. (2013). Neural processes underlying the “same”-“different” judgment of two simultaneously presented objects-an EEG study. *PLoS One*, **8**, e81737.

Zhou, X. and Mondloch, C.J. (2016). Recognizing ‘‘Bella Swan’’ and ‘‘Hermione Granger’’: no own-race advantage in recognizing photos of famous faces. *Perception*, **45**, 1426-1429.

Appendix 5.1

Matrix summarising how participant (*n* = 22) unconstrained choice of groups changed when selection was constrained to four. “Free” shows how many groups each participant sorted the 80 images into when choice was unconstrained. “Constrained” shows how “Free” group images were sorted under constrained conditions, by both number and percentage. The final column shows how many new groups the images in each unconstrained group was resorted into, i.e. one means that image choice was consistent as images in the “Free” group were included in the *same* group when resorted.

Free		**Constrained**								**No. groups from original**
		(No.)				(%)				
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	
1	**1**	1	0	0	19	0.05	0.00	0.00	0.95	2
	2	1	6	5	1	0.08	0.46	0.38	0.08	4
	3	0	18	6	1	0.00	0.72	0.24	0.04	3
	4	8	2	5	7	0.36	0.09	0.23	0.32	4
		10	**26**	**16**	**28**					
2	**1**	15	1	0	0	0.94	0.06	0.00	0.00	2
	2	0	19	0	0	0.00	1.00	0.00	0.00	1
	3	0	0	15	0	0.00	0.00	1.00	0.00	1
	4	0	8	1	7	0.00	0.50	0.06	0.44	3
	5	1	0	0	13	0.07	0.00	0.00	0.93	2
		16	**28**	**16**	**20**					
3	**1**	0	17	10	0	0.00	0.63	0.37	0.00	2
	2	12	0	1	2	0.80	0.00	0.07	0.13	3
	3	0	0	9	0	0.00	0.00	1.00	0.00	1
	4	0	0	2	10	0.00	0.00	0.17	0.83	2
	5	7	0	5	5	0.41	0.00	0.29	0.29	3
		19	**17**	**27**	**17**					
4	**1**	3	0	0	1	0.75	0.00	0.00	0.25	2
	2	3	3	0	3	0.33	0.33	0.00	0.33	3
	3	1	0	1	2	0.25	0.00	0.25	0.50	3
	4	8	5	3	2	0.44	0.28	0.17	0.11	4

	5	3	1	1	1	0.50	0.17	0.17	0.17	4
	6	5	2	3	8	0.28	0.11	0.17	0.44	4
	7	1	5	7	2	0.07	0.33	0.47	0.13	4
	8	3	2	0	1	0.50	0.33	0.00	0.17	3
		27	**18**	**15**	**20**					
5	**1**	0	0	1	4	0.00	0.00	0.20	0.80	2
	2	0	0	1	1	0.00	0.00	0.50	0.50	2
	3	4	0	1	0	0.80	0.00	0.20	0.00	2
	4	3	1	1	2	0.43	0.14	0.14	0.29	4
	5	0	6	1	0	0.00	0.86	0.14	0.00	2
	6	1	0	0	8	0.11	0.00	0.00	0.89	2
	7	0	0	3	0	0.00	0.00	1.00	0.00	1
	8	11	1	0	1	0.85	0.08	0.00	0.08	3
	9	0	6	2	1	0.00	0.67	0.22	0.11	3
	10	2	2	5	0	0.22	0.22	0.56	0.00	3
	11	1	0	5	5	0.09	0.00	0.45	0.45	3
		22	**16**	**20**	**22**					
6	**1**	3	12	3	1	0.16	0.63	0.16	0.05	4
	2	1	0	0	3	0.25	0.00	0.00	0.75	2
	3	2	0	1	1	0.50	0.00	0.25	0.25	3
	4	3	2	4	2	0.27	0.18	0.36	0.18	4
	5	2	1	0	10	0.15	0.08	0.00	0.77	3
	6	0	0	1	3	0.00	0.00	0.25	0.75	2
	7	1	1	0	5	0.14	0.14	0.00	0.71	3
	8	5	1	0	0	0.83	0.17	0.00	0.00	2
	9	11	1	0	0	0.92	0.08	0.00	0.00	2
		28	**18**	**9**	**25**					
7	**1**	0	0	1	0	0.00	0.00	1.00	0.00	1
	2	1	2	3	0	0.17	0.33	0.50	0.00	3
	3	19	0	0	0	1.00	0.00	0.00	0.00	1
	4	0	5	18	0	0.00	0.22	0.78	0.00	2
	5	1	0	0	2	0.33	0.00	0.00	0.67	2
	6	3	0	2	10	0.20	0.00	0.13	0.67	3
	7	0	0	0	2	0.00	0.00	0.00	1.00	1
	8	0	1	2	2	0.00	0.20	0.40	0.40	3
	9	0	6	0	0	0.00	1.00	0.00	0.00	1
		24	**14**	**26**	**16**					
8	**1**	0	6	0	0	0.00	1.00	0.00	0.00	1
	2	0	7	0	0	0.00	1.00	0.00	0.00	1
	3	9	0	0	0	1.00	0.00	0.00	0.00	1
	4	0	7	0	0	0.00	1.00	0.00	0.00	1

	5	0	0	15	0	0.00	0.00	1.00	0.00	1
	6	0	0	1	0	0.00	0.00	1.00	0.00	1
	7	7	0	0	0	1.00	0.00	0.00	0.00	1
	8	0	0	0	12	0.00	0.00	0.00	1.00	1
	9	16	0	0	0	1.00	0.00	0.00	0.00	1
		32	**20**	**16**	**12**					
9	**1**	1	0	10	0	0.09	0.00	0.91	0.00	2
	2	3	1	1	0	0.60	0.20	0.20	0.00	3
	3	0	0	0	18	0.00	0.00	0.00	1.00	1
	4	1	19	0	1	0.05	0.90	0.00	0.05	3
	5	20	0	5	0	0.80	0.00	0.20	0.00	2
		25	**20**	**16**	**19**					
10	**1**	0	6	5	0	0.00	0.55	0.45	0.00	2
	2	15	0	1	1	0.88	0.00	0.06	0.06	3
	3	0	2	0	0	0.00	1.00	0.00	0.00	1
	4	0	0	3	2	0.00	0.00	0.60	0.40	2
	5	1	3	1	0	0.20	0.60	0.20	0.00	3
	6	0	0	0	7	0.00	0.00	0.00	1.00	1
	7	1	0	2	2	0.20	0.00	0.40	0.40	3
	8	0	1	0	7	0.00	0.13	0.00	0.88	2
	9	0	0	0	5	0.00	0.00	0.00	1.00	1
	10	0	1	0	10	0.00	0.09	0.00	0.91	2
	11	0	0	1	1	0.00	0.00	0.50	0.50	2
	12	0	1	0	0	0.00	1.00	0.00	0.00	1
	13	0	0	1	0	0.00	0.00	1.00	0.00	1
		17	**14**	**14**	**35**					
11	**1**	3	13	0	0	0.19	0.81	0.00	0.00	2
	2	0	0	6	3	0.00	0.00	0.67	0.33	2
	3	0	0	13	3	0.00	0.00	0.81	0.19	2
	4	20	6	0	13	0.51	0.15	0.00	0.33	3
		23	**19**	**19**	**19**					
12	**1**	13	2	1	1	0.76	0.12	0.06	0.06	4
	2	13	1	2	4	0.65	0.05	0.10	0.20	4
	3	5	2	0	6	0.38	0.15	0.00	0.46	3
	4	3	4	4	6	0.18	0.24	0.24	0.35	4
	5	4	5	3	1	0.31	0.38	0.23	0.08	4
		38	**14**	**10**	**18**					
13	**1**	2	2	2	16	0.09	0.09	0.09	0.73	4
	2	1	1	11	0	0.08	0.08	0.85	0.00	3
	3	1	0	0	11	0.08	0.00	0.00	0.92	2

	4	3	16	0	1	0.15	0.80	0.00	0.05	3
	5	12	0	0	1	0.92	0.00	0.00	0.08	2
		19	**19**	**13**	**29**					
14	**1**	20	0	0	0	1.00	0.00	0.00	0.00	1
	2	4	0	12	0	0.25	0.00	0.75	0.00	2
	3	0	3	0	16	0.00	0.16	0.00	0.84	2
	4	0	4	4	0	0.00	0.50	0.50	0.00	2
	5	0	11	3	3	0.00	0.65	0.18	0.18	3
		24	**18**	**19**	**19**					
15	**1**	17	0	0	0	1.00	0.00	0.00	0.00	1
	2	1	1	0	0	0.50	0.50	0.00	0.00	2
	3	0	5	1	0	0.00	0.83	0.17	0.00	2
	4	1	0	2	0	0.33	0.00	0.67	0.00	2
	5	0	0	10	0	0.00	0.00	1.00	0.00	1
	6	0	0	1	0	0.00	0.00	1.00	0.00	1
	7	0	20	0	0	0.00	1.00	0.00	0.00	1
	8	0	0	0	4	0.00	0.00	0.00	1.00	1
	9	0	4	0	13	0.00	0.24	0.00	0.76	2
		19	**30**	**14**	**17**					
16	**1**	17	0	2	0	0.89	0.00	0.11	0.00	2
	2	0	0	0	18	0.00	0.00	0.00	1.00	1
	3	2	0	21	0	0.09	0.00	0.91	0.00	2
	4	0	19	0	1	0.00	0.95	0.00	0.05	2
		19	**19**	**23**	**19**					
17	**1**	0	0	13	0	0.00	0.00	1.00	0.00	1
	2	0	0	0	11	0.00	0.00	0.00	1.00	1
	3	0	7	0	1	0.00	0.88	0.00	0.13	2
	4	0	7	0	0	0.00	1.00	0.00	0.00	1
	5	0	7	0	0	0.00	1.00	0.00	0.00	1
	6	0	1	0	6	0.00	0.14	0.00	0.86	2
	7	0	4	0	1	0.00	0.80	0.00	0.20	2
	8	19	0	0	3	0.86	0.00	0.00	0.14	2
		19	**26**	**13**	**22**					
18	**1**	0	0	0	5	0.00	0.00	0.00	1.00	1
	2	0	0	0	9	0.00	0.00	0.00	1.00	1
	3	0	0	1	8	0.00	0.00	0.11	0.89	2
	4	0	0	20	1	0.00	0.00	0.95	0.05	2
	5	0	16	0	0	0.00	1.00	0.00	0.00	1
	6	20	0	0	0	1.00	0.00	0.00	0.00	1
		20	**16**	**21**	**23**					

19	**1**	11	2	0	1	0.79	0.14	0.00	0.07	3
	2	0	13	1	0	0.00	0.93	0.07	0.00	2
	3	0	2	7	1	0.00	0.20	0.70	0.10	3
	4	0	0	0	17	0.00	0.00	0.00	1.00	1
	5	23	0	2	0	0.92	0.00	0.08	0.00	2
		34	**17**	**10**	**19**					
20	**1**	0	12	0	1	0.00	0.92	0.00	0.08	2
	2	0	14	1	0	0.00	0.93	0.07	0.00	2
	3	5	0	2	0	0.71	0.00	0.29	0.00	2
	4	0	0	12	0	0.00	0.00	1.00	0.00	1
	5	15	0	2	0	0.88	0.00	0.12	0.00	2
	6	0	3	0	13	0.00	0.19	0.00	0.81	2
		20	**29**	**17**	**14**					
21	**1**	2	12	0	2	0.13	0.75	0.00	0.13	3
	2	5	3	0	20	0.18	0.11	0.00	0.71	3
	3	2	0	0	6	0.25	0.00	0.00	0.75	2
	4	10	0	0	0	1.00	0.00	0.00	0.00	1
	5	1	0	4	0	0.20	0.00	0.80	0.00	2
	6	0	0	11	2	0.00	0.00	0.85	0.15	2
		20	**15**	**15**	**30**					
22	**1**	0	0	18	0	0.00	0.00	1.00	0.00	1
	2	0	6	0	0	0.00	1.00	0.00	0.00	1
	3	0	13	0	1	0.00	0.93	0.00	0.07	2
	4	0	9	0	3	0.00	0.75	0.00	0.25	2
	5	20	0	0	0	1.00	0.00	0.00	0.00	1
	6	0	0	0	10	0.00	0.00	0.00	1.00	1
		20	**28**	**18**	**14**					

Appendix 5.2

Participant___________

Many thanks helping with this survey. The survey is anonymous, but we would like to gather some information about you.

Do you agree to participate in this survey? ○ Yes ○ No

1. What is your gender?

○ Female ○ Male ○ Other

2. What is your age?

3. Do you consider yourself to have normal vision?

○ Yes ○ No, but my vision is corrected to normal with glasses or contact lenses ○ No, I have a visual impairment

4. Have you ever seen a newt?

○ Yes Do you know what species? ○ No

5. Have you ever handled a newt?

○ Yes Do you know what species? ○ No

6. How do you rate your ability at identifying species?

○ Very bad ○ Bad ○ Neither good nor bad ○ Good ○ Very good

7. How do you rate your ability at identifying amphibians?

○ Very bad ○ Bad ○ Neither good nor bad ○ Good ○ Very good

Thank you very much for answering these questions and participating in the survey.

Giraffe or prate? @teenybiscuit

Chapter 6.

6 General Discussion

This thesis aimed to investigate patterns and processes in species identification. As the first in-depth study of its kind, methods widely used in forensic face recognition were applied to images of species. Not only do the findings in this thesis demonstrate parallels with the outcomes of face matching tasks, but also show that there is more to learn from this area of psychology. Each task showed that based on visual cues alone, accuracy and consistency in species identification varied both *within* (i.e. inconsistent decisions on the *same* stimuli in chapter 2) and *between* participants (i.e. a range of judgements in all tasks). There was also a lack of uniformity across species for each task. Furthermore, this variation in accuracy was further reflected across all levels of experience and expertise with the model taxa, whether stimuli were presented under optimised conditions (i.e. simultaneously as pairs without context in chapters 2 and 3), or as ambient images (i.e. non-manipulated images presented concurrently in chapters 4 and 5). This evidence shows that this interdisciplinary approach has been informative, and that further investigation into these differences is warranted.

6.1 Contributions to knowledge and applications

It is widely acknowledged that high quality data is critical for effective decision-making, yet the accuracy of species observations is rarely discussed. Despite the widespread use of species data, individual ability to accurately identify species is generally unknown, untested and assumed, rather than based on proven performance. The tasks conducted in this thesis showed variation in individual ability of both experts and non-experts, that accuracy is independent of experience, and that experts do not always agree. These findings demonstrate a need for identification abilities to be ascertained for monitoring purposes, and encompassed in discussion about data bias and error. Some expert participants noted that experienced observers are aware of their own limitations, yet few studies draw attention to the perceived or tested competency of those collecting species data, even though these data inform important decisions in research, policy, and planning. There is also little mention of *how* identification is made. For example, do experienced observers rely on prior knowledge, or do they refer to identification guides? If identification aids have been consulted, naming them could help with repeatability. Have those data have been validated, and by whom? Species data is collected in a way that suits the needs of the recorder (nbn.org.uk), and there are a variety of databases to which these records can be added (see section 1.5.1). These data may already be verified (ensuring the accuracy of the identification) or validated (probability of the record being correct), but these are not necessarily requirements for submission (see nbn.org.uk for case studies). Comprehensive metadata for individual datasets may help the end user identify any perceived knowledge gaps in that data.

While bias in biological recording is well known (Isaac & Pocock 2015), error in identification accuracy is rarely discussed. For example, when authors have collected data on species observations, have their records been validated by another person? Equally, some monitoring programmes incorporate data from different sources (Ruiz-Gutierrez et al. 2016), but if these data are used in a scientific study, have the authors made an *a priori* assumption that species observations are 100% accurate? Furthermore, there is little mention of acceptable levels of accuracy or margins of error. When research is conducted using data obtained from publicly accessible datasets, metadata provided with datasets may provide end users with information that could help them to ascertain a level of confidence in the data they download, but some large datasets are from numerous sources (e.g. NBN Gateway). Given that biological records provide data for research in a number of biological disciplines (Powney & Isaac 2015), as well as modelling to fill data gaps and model species across spatial or temporal scales (Evans et al. 2016; Joppa et al. 2016), a discussion around possible identification error in the data would be prudent. This should, however, be considered on a case-by-case basis, as accuracy is not uniform across participants or taxa. These discussions should also remain fluid, as rapid advancements in technology mean that methods of species identification are likely to develop in different, and perhaps unpredictable ways.

Historically, species identification would have been conducted by a small number of specialists, and taken place in the field or by collecting specimens. However improvements in technology have allowed a wider audience to participate in data to participate in data collection, and collection, and provide large amount of data, especially as images (Pimm et al. 2015). However, this thesis shows that data collected in this manner is not without its own limitations. Whether photographs, satellite imagery, or footage from devices such as camera phones or drones, images captured in this manner can be highly variable (Jenkins et al. 2011). The identification process is often a binary choice of an unknown specimen being one species *or* another. However, in practice, judgements are more akin to a Bayesian ‘more alike’ or ‘less alike’ process, and perhaps observations should include confidence levels (see Wilkinson & Evans 2009). Chapter 2 showed that experts are more likely to acknowledge uncertainty in their decisions, as communicated by participants during the study, as well as in replies to ‘tweets’ of the papers from this chapter (Austen et al. 2016). For example, one participant noted “*I have answered 'don't know' for quite a lot of pairings - perhaps unhelpfully. This reflects my personal approach to ID caution and experience of plasticity, which may vary between 'experts'. Perhaps there should be one or two more options to split the 'don't knows' into e.g. the 'insufficient knowledge don't knows' and the 'knowledgeable don't knows' who appreciate both possibilities of same and different species*” (pers. comm. 24/7/2014). However, such a cautious approach raises the issue of how to treat disputed records. Discounting observations may appear to be a waste of data, but the inclusion of ambiguous observations may lead to the wrong conclusions (Solow et al. 2012). An acceptable rate of identification error depends on the aims of the research.

The level of expected or accepted error in species identification should be considered during the design phase of any study involving observation data. In forensic face matching, the security aspect makes accuracy of great importance, and misidentification can have serious repercussions (Jenkins and Burton 2008b). However, although a high level of accuracy is preferable in most scenarios, it may not be paramount in species identification and situation dependent. Face matching literature notes that human identification is fallible, and compounded by people thinking that they are face experts (see Johnston & Edmonds 2009). Conversely, as noted above, some species recorders are aware of their limitations. How species data will be used should provide an indication of acceptable levels of error. For example, chapter 4 showed a notable lack of agreement for *Lissotriton helveticus* and *L. vulgaris*, which is important if data collected are used to determine populations and distributions, especially for modelling purposes (Costa et al. 2015). However, if surveying for development work, neither would trigger mitigation measures and any confusion would not be problematic for the outcomes of the survey. Conversely, observing *any* bat flying out of a building will be important if surveying for development as all bats (Order: Chiroptera) are fully protected under UK law (www.gov.uk, 2017). However, such a general observation may not be informative if species numbers and distributions are being determined. There is also the consideration of whether the observer is *expected* to possess good identification skills.

Chapters 2 and 4 indicated that despite possible expectations, experience is no guarantee of accuracy when identifying species from images. It is also irrespective of self-perceived identification ability. This mirrors findings from face recognition studies where neither forensic police officers (Burton et al. 1999) nor passport officers (White et al. 2014) were more accurate than other participants for unfamiliar faces. Perhaps the question is *where* the expertise lies. The similarity in identification rates between experts and non-experts in chapter 2 should not detract from skills that the expert participants have. Their species knowledge may lie with knowing the habitats favoured by certain bumblebees, their food sources, mating habits, etc. Unlike the face matching research, participant familiarity with bumblebee species did not match the same species using different illustrations, but it may be that these specialists are so used to handling specimens, that images do not provide the cues that they rely on. As with the coining of the term ‘super-recognisers’ and their recruitment to forensic face matching roles (Russell, Duchaine & Nakayama 2009; Bobak, Hancock & Bate 2016), it could be that the expertise lies in the ability to accurately match different images of the same individual or species. The individual ability to consistently identify unfamiliar faces with high accuracy has led to the Metropolitan Police recruiting super-recognisers from their ranks (Robertson et al. 2016). In accordance with this application of skills, perhaps the ability to identify whether species are the same or different is independent of a specialist knowledge with the model taxa, as seen in chapters 3 and 5. One study found that a small group of face image experts outperformed the public (Wilkinson & Evans 2009), but these results compared averages from each group, whereas this thesis show that individual scores are much more informative.

This research also shows that while experience can provide a healthy level of doubt when naming species, accurate species identification is not solely the domain of experts. Moreover, the results in chapter 4 indicate that perception of identification skills may not match actual ability, or be the same in every situation (e.g. differences in accuracy when identifying specimens rather than images). These are important consideration when using data to inform policy. Many policies that affect the environment, such as those related to pollution control, agriculture, fisheries and wildlife, originate in European Union (EU) legislation. With the proposed changes to EU membership, this legislation will be reviewed and decisions on whether to retain, amend, or abandon these policies will be influenced by evidence made available to policy makers. Data and statistics are frequently used to both direct and object to policy decisions, and are sometimes sought within a short time period. It is imperative that any new or reformed policies are based on robust, reliable information, whether from professional or amateur datasets.

6.1.1 Training in species identification

Whether heuristically or as part of a structured programme, the process by which people learn to identify species is likely to involve training. Training courses vary in their delivery and outcomes, but their effectiveness is sometimes measured by a test on completion. However, learning can also be unintentional and happen even when little attention is being paid (Shanks 2005), and participants on a training course may be unaware of their own abilities. Testing participants before *and* after training may be more informative in assessing improvements in the skills of individuals, as well as the efficacy of course. Learning about species generally involves more than just visual identification as tested in this thesis, for example typical habitat, distribution, and behaviour. However, training needs to be appropriate to participant ability and objectives, and training material should reflect this. For example, an image under optimised conditions does not represent field conditions, but when used to identify an isolated specimen (e.g. a bumblebee caught temporarily in a tube and placed against white paper), this type of image may be an effective training tool. Although identification by this process is restricted, images with context can prove highly informative. However, this additional information can also bias decisions on identification process, and a correct identification may be discounted if the image background is not typically associated with that species (Gibbon, Bindemann & Roberts 2015).

With regards to general identification guides and keys, a certain level of expertise is required to write them, but how accessible they are to those with less experience is not always clear. A recent Twitter entry reads “Test the key… works well. Mentor suggests we test with ‘intelligent ignoramus’ – someone who doesn’t already know what it is. Key fails.” (Matt Keevil, Twitter 19/5/17). This draws some parallels with industry, where a lack of usability training means that some people are being asked to use or advise on products which they do not fully understand (Rubin and Chisnell 2008), Part of the learning process for any novice is to discover which guides work for them. Wider accessibility to electronic guides, plus the number of online platforms and websites related to species identification, provide more information but do not necessarily improve understanding. However, the Internet allows recommendations and reviews on identification guides to be sought, and uncertainty in identification discussed widely.

6.1.2 Undetected factors and influences

The tasks in this thesis required participants to make same-different judgements on stimuli presented under different conditions. In chapters 2 and 3, model taxa had the same body plan, a limited range of colours, and variation in stripes (if at all), and were presented whole, without context and from the same perspective. The taxa in chapters 4 and 5 were presented as both whole and part organisms, in context, and from a range of perspectives. As every task demonstrated variation in judgements as to whether images were of the same species, the extent to which shape and colour are involved in decision-making processes warrants further investigation. Research into same-different judgements on shape and colour found that for 16 animals (including pineapple, orange, avocado, grapes, and peppers) and 16 fruits and vegetables (including dog, rabbit, shark, lion, and giraffe), identification became more difficult when shape was altered rather than colour (Scorolli & Borghi 2015). Further studies with manipulated images of organisms (e.g. outline only, greyscale, or incorrect colouration) could further investigate the influence of colour over shape. Also, the perception of colour against varying backgrounds, or under different lighting conditions is worth exploring.

There are also other cues that may influence species identification. Although vision is the primary mode of percept (Kaspar 2013), visual species identification may be influenced by the simultaneous presentation of other stimuli, for example sound or odour. A recent study showed that adolescents with impaired hearing were better at visual processing than their counterparts with unimpaired hearing (Megreya & Bindemann 2017). Decisions may also be influenced by ambient or defined sounds, and odours (Wilson & Stevenson 2003). A task comparing accuracy between purely visual and multisensory identification could explore this, testing both match (same species) and mismatch (different species) visual and auditory cues. Also, the influence of seemingly peripheral cues, such as image context (e.g. habitat or other species) or information provided by identification guides is worth exploring.

6.1.3 The economics of accurate identification

As previously discussed, accuracy in species identification can have far reaching consequences, with misidentification potentially affecting human well-being and the economy. Training or recruiting people in species identification will involve some expenditure, and while this may not be a priority in budgetary terms, this lack of investment could prove costly. In this context, the expense of targeted, effective training should be balanced with the socio-economic costs of *not* training identification abilities. Moreover, this should not be decided in relation to whether species observations are made by volunteers or professionals, but how the data will be applied and which decisions it will inform. For example, some citizen science projects are mainly aimed at public participation and benefit local communities, whereas as others collect data which show trends in species numbers and are used as evidence (e.g. managements plans and policy). Conversely, data collected by professional ecologists may only ever appear in a client's report and nowhere else. Given the variability in identification accuracy both between individuals and species, the findings in this thesis suggest that in addition to training, identification skills should be tested. Moreover, training and testing should be targeted, conducted with a variety of taxa, and should include more than one image of the same species or individual. For example, training for a crowd-sourced camera trap study on individuals within a population will differ from that required by customs officials looking for CITES species. In conservation science, expert knowledge is widely used, but its reliability depends on the rigour with which it is acquired (Martin et al. 2012). There are organisations that draw together expertise to resolve environmental problems and evaluate the outcomes of environmental decisions. These collectives, such as the Centre of Excellence for Biosecurity Risk Analysis (http://cebra.unimelb.edu.au) and the ARC Centre of Excellence for Environmental Decisions (http://ceed.edu.au), advocate the optimal use of information and data. Likewise, the accuracy and effectiveness of species data used as evidence in management, planning, and policy should be assessed and evaluated.

6.2 Practical considerations and limitations

The application of methods from face recognition studies provided novel insights, as well as challenges similar to those faced in that research. Testing accuracy in species identification also requires the correct balance between controlled laboratory and field conditions. Field observations are made on organisms that may be moving, possibly partially obscured, sometimes with time constraints, and for species that exhibit different characteristics at different life stages. However, to include all of these factors in the study design would make results difficult to interpret. Species are identified under a range of condition, for example as live specimens, in museum collections, or from images and video footage. Nonetheless, the findings in this thesis show that on visual cues *alone*, accurate identification is highly variable and non-uniform.

As with any study that recruits volunteers, recruitment of participant warrants discussion. As the tasks were based on visual cues, there was no *a priori* reason to select volunteers on anything but having normal, or corrected-to-normal vision. For expert participation in chapters 2 and 4, volunteers were recruited via specialist conservation organisations. Non-expert volunteers were mainly recruited via DICE, and in chapter 3, the majority of volunteers were from the School of Psychology, where students must earn course credits by participating in a study of their choice. Unless there is a sufficient budget to pay a company to recruit participants, survey volunteers are likely to revolve around the network of those delivering the tasks. Motivations for participation are personal, examples being making a difference or a positive contribution to society (Hobbs & White 2012). Although chapter 2 found that non-experts interested in biodiversity and conservation did not outperform other non-experts, links between positive attitudes towards the natural world, identification and confidence in decisions warrant further research. This thesis has discussed the contribution of amateur naturalists in terms of knowledge and recording, but this is only part of the story. Support for nature conservation can be found in membership of local, national and international organisations that aim to protect wildlife in general (e.g. Wildlife Trusts, World Wildlife Fund) and specific taxon groups (e.g. BirdLife

International, Plantlife). Furthermore, the human benefits of interacting with the natural world can be measured, such as the Connectedness to Nature Scale (Mayer and Frantz 2004), and this may be another motivation that affects accurate identification.

6.2.1 Technology

Emerging technologies have revolutionised how we view the world, creating new insights and new challenges (Pimm et al. 2015). The increased use of remote sensing has resulted in volumes of images that require processing, and this has given rise to initiatives through which the identification of species is being aided by crowdsourcing (Silvertown et al. 2015; Swanson et al. 2015; Daume & Galaz 2016; Chandler et al. 2017). This approach helps plug knowledge gaps, but has limitations (Verma, van der Wal & Fischer 2016) as shown in chapters 4 and 5. Although images do not necessarily offer the same information as seeing an organism *in-situ*, digital images can be stored, shared and viewed equally by all. For example, an image can visually support an observation, can be shared among experts for verification, and can be retrieved at a later date, unlike verbal or written description, which may differ between individuals describing the same observation. Moreover, technology can also be used to improve identification accuracy by making images more general. For example, face matching research shows that accuracy improves when several images of an individual are merged to create an ‘average’ image (Jenkins & Burton 2011; Taubert, Weldon & Parr 2016). As chapter 2 shows, images of the same species from different guides vary enough to appear to be different species, the creation of an ‘average’ image of a species may help with identification. There is also the option of distorting a single image to simulate possible variations (Macleod, O’Neill & Walsh 2007).

Following advances in technology, there has also been a call for automated identification (Gaston & O’Neill 2004; MacLeod, Benfield & Culverhouse 2010). Although an understanding of the mechanisms behind the visual processing may aid automated identification, it is unlikely to depend upon it. Work on automated taxon identification is based on objectifying or transcribing an organism into a shape that can be subjected to mathematical analysis (e.g. Fourier). Examples of automatic identification in non-human species include plants (Barré et al. 2017), plankton (Benfield et al. 2007), and wasps (MacLeod, O’Neill & Walsh 2007). However, despite the increasing use of technology for the purposes of security, accuracy of automated face recognition devices remains poor (Burton & Jenkins 2011). Neither humans nor machines are good at matching photographs to a face (Jenkins & Burton 2008; Bindemann & Burton 2009), although recent literature found that computers are superior at matching frontal photographs, and humans are superior at more difficult images (Phillips & O’Toole 2014). Accuracy in automatic identification is likely to be a result of improving algorithms and large training sets rather than an understanding of visual cues in identification. Although conspecifics can vary greatly in appearance, this technology may be able to help with the identification of individual animals.

6.3 Further applications of psychology in understanding species identification

The application of methods from forensic face recognition studies has been highly informative to understanding accuracy in species identification. This thesis has only applied a few methods, and there are more lessons to be learned from this area of research. For example, one widely used method that could provide a wealth of information is the use of eye tracking equipment. This technology explores the observer’s eye movements and provides information regarding the mechanisms behind visual searches (Duchowski 2007; Eckstein 2011). Eye tracking is successfully used in other domains, for example how radiologists with different levels of experience view chest images (Fig. 6.1) (Manning et al. 2006). This study found that experienced and inexperienced observers had distinctly different search strategies. As well as being of general interest to the process of species identification, repeating chapter 2 using eye-tracking technology could investigate whether observers with different expertise showed patterns in their visual coverage. The simultaneous collection of qualitative data in such an experiment may also prove useful. A comparison between what observers *thought* they were looking at (e.g. shape, body composition, colour, markings etc.), and where they were *actually* looking could provide information on the cognitive process of visual identification.

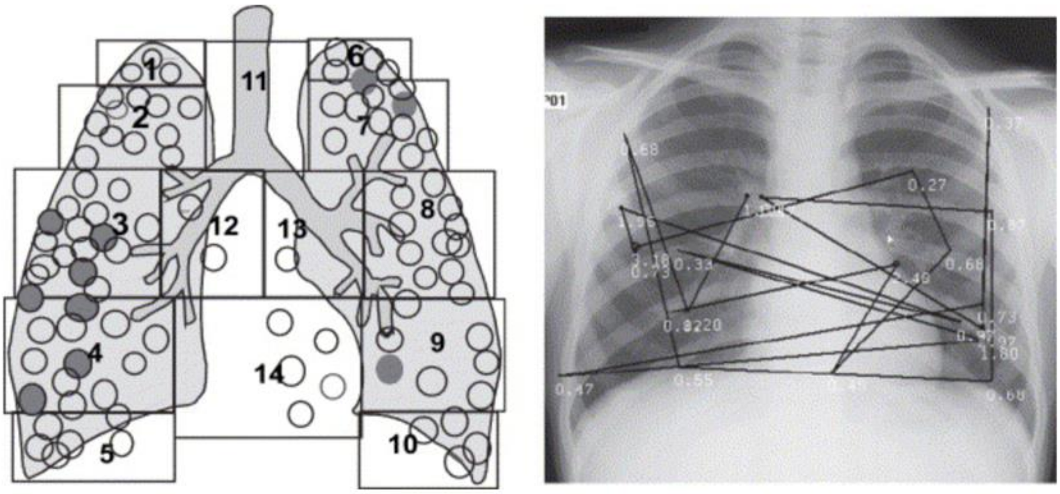


Figure 6.1 An example of how eye tracking is used in medical training investigated how observers with different levels of experience searched a chest image: (*left*) images were divided into 14 zones with nodules of interest (shaded discs); (*right*) shows that experienced radiologists do not examine large areas of the image. (From Manning et al. 2006)

6.4 Closing remarks

Species are crucial for studying a range of subjects, for example ecology, biodiversity, evolution, agriculture and evolution, as well as used for dating strata (Hohenegger 2014). Furthermore, species identification has a large part to play in society, and is especially important to those for whom understanding of nature is inherent to their livelihoods (Shipman & Boster 2008). An understanding of species numbers, distributions and interactions are imperative in many biological disciplines, this information needs to be as accurate as possible. The research in this thesis has drawn on different disciplines and provided a different perspective to investigating the process of species identification. Interdisciplinary research in academia has become more prevalent in recent years (Jacobs 2014), but the amalgamation of knowledge from different disciplines should not be restricted to the perimeters of academic institutions. The use of methods from face recognition research has produced findings that may aid training courses run by NGOs, and the application of other social science research may help at local levels. For example, decisions on managing nature reserves could benefit from understanding who visits the reserve, their motivations, and their perceived benefits of doing so (Keniger et al. 2013).

This thesis found that species identification is not the domain of experts, experience is no guarantee of accuracy, and that intra-specific and inter-specific variation is not uniform. These outcomes draw many parallels with face recognition research, and suggest that novel approaches to existing questions can further our understanding of patterns and processes in the natural world. Environmental problems are responsible for some of the greatest challenges faced by humans in the world today (Dillon 2016). Biodiversity loss, food insecurity, the impact of invasive species and the effects of climate change, as well as these factors combined, have been deemed as ‘wicked’ problems, which are difficult to comprehend, define or resolve (Termeer, Dewulf & Breeman 2013; Strassheim 2015; Dillon 2016). Addressing and resolving such problems requires input from a range of specialists (Ellwood et al. 2016), not only to break these elements down into workable components, but also to research, deliver and communicate possible solutions. The hope is that this thesis provides information that could help ameliorate this process.

6.5 References

Austen, G.E., Bindemann, M. and Roberts D.L. (2016). Species identification by experts and non-experts: comparing images from field guides. *Scientific Reports* **6**, 33634.

Barré, P., Stöver, B.C., Müller, K.F. and Steinhage, V. (2017). LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics*, **40**, 50–56.

Benfield, M., Grosjean, P., Culverhouse, P.F., Irigoien, X., Sieracki, M.E., Lopez-Urrutia, A., Dam, H.G., Hu, Q., Davis, C.S., Hansen, A. and Pilskaln, C.H. (2007). RAPID: Research on automated plankton identification. *Oceanography*, **20**, 172–187.

Bindemann, M. and Burton, A.M. (2009). The role of color in human face detection. *Cognitive Science*, **33**, 1144-1156.

Bobak, A.K., Hancock, P.J.B. and Bate, S. (2016). Super-recognisers in action: evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, **30**, 81–91.

Burton, A.M. and Jenkins, R. (2011). Unfamiliar face perception. In: Calder, A.J., Rhodes, G., Johnson, M.H. and Haxby, J.V. eds. *The Oxford Handbook of Face Perception.* Oxford University Press, Oxford, 303.

Burton, A.M., Wilson, S., Cowan, M. and Bruce, V. (1999). Face recognition in poor-quality video: evidence from security surveillance. *Psychological Science*, **10**, 243–248.

Chandler, M. See, L., Buesching, C.D., Cousins, J.A., Gillies, C., Kays, R.W., Newman, C., Pereira, H.M. and Tiago, P. (2017). Involving Citizen Scientists in Biodiversity Observation. In: *The GEO Handbook on Biodiversity Observation Networks*. Cham: Springer International Publishing, pp. 211–237.

Costa, H., Foody, G.M., Jiménez, S. and Silva, L., (2015). Impacts of species misidentification on species distribution modeling with presence-only data. *ISPRS International Journal of Geo-Information*, **4**, 2496-2518.

Daume, S. and Galaz, V. (2016). “Anyone know what species this is?” - Twitter conversations as embryonic citizen science communities. *PLoS One*, **11**, e0151387.

Dillon, J. (2016). 50 Years of JBE: from science and environmental education to civic science. *Journal of Biological Education*, **50**, 120-122.

Duchowski, A.T. (2007). *Eye Tracking Methodology : Theory and Practice*. Springer, London.

Eckstein, M.P. (2011). Visual search: a retrospective. *Journal of Vision*, **11**, 14–14.

Ellwood, E.R., Crimmins, T.M. and Miller-Rushing, A.J. (2017). Citizen science and conservation: recommendations for a rapidly moving field. *Biological Conservation*.

Evans, M., Merow, C., Record, S., McMahon, S.M. and Enquist, B.J. (2016). Towards process-based range modeling of many species. *Trends in Ecology & Evolution*, **31**, 860-871.

Gaston, K.J. and O'Neill, M.A. (2004). Automated species identification: why not? *Philosophical transactions of the Royal Society B,* **359**, 655-667.

Gibbon, G.E.M., Bindemann, M. and Roberts, D.L. (2015). Factors affecting the identification of individual mountain bongo antelope. *PeerJ*, **3**, e1303.

HM Government *Bats: Protection and Licences* [Online]. Available from: https://www.gov.uk/guidance/bats-protection-surveys-and-licences [Accessed: 5 June 2017].

Hobbs, S.J. and White, P.C.L. (2012). Motivations and barriers in relation to community participation in biodiversity recording. *Journal for Nature Conservation*, **20**, 364–373.

Hohenegger, J. (2014). Species as the basic units in evolution and biodiversity: recognition of species in the recent and geological past as exemplified by larger foraminifera. *Gondwana Research*, **25**, 707–728.

Isaac, N.J.B. and Pocock, M.J.O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, **115**, 522-531.

Jacobs, J.A. (2014). *In Defense of Disciplines: Interdisciplinarity and Specialization in the Research University*. University of Chicago Press, Chicago.

Jenkins, R. and Burton, A.M. (2008). 100% accuracy in automatic face recognition. *Science*, **319**, 435.

Jenkins, R. and Burton, A. (2008b). Limitations in facial identification: the evidence. *Justice of the Peace*, **172**, 4-6.

Jenkins, R. and Burton, A.M. (2011). Stable face representations. *Philosophical transactions of the Royal Society B*, **366**, 1671-1683.

Jenkins, R. White, D., Van Montfort, X. and Burton, A.M. (2011). Variability in photos of the same face. *Cognition*, **121**, 313-323.

Johnston, R.A. and Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: a review. *Memory*, **17**, 577-596.

Joppa, L.N., O'Connor, B., Visconti, P., Smith, C., Geldmann, J., Hoffmann, M., Watson, J.E., Butchart, S.H., Virah-Sawmy, M., Halpern, B.S. and Ahmed, S.E., (2016). Filling in biodiversity threat gaps. *Science*, **352,** 416–418.

Kaspar, K. (2013). What guides visual overt attention under natural conditions? Past and future research. *ISRN neuroscience*.

Keniger, L.E., Gaston, K.J., Irvine, K.N. and Fuller, R.A (2013). What are the benefits of interacting with nature? *International Journal of Environmental Research and Public Health*, **10**, 913-935.

MacLeod, N., Benfield, M. and Culverhouse, P. (2010). Time to automate identification. *Nature*, **467**, 154-155.

MacLeod, N., O'Neill, M. and Walsh, A. (2007). Automated tools for the identification of taxa from morphological data: face recognition in wasps. In: MacLeod, N. ed. *Automated taxon recognition in systematics: theory, approaches and applications.* CRC Press, Boca Raton, 153-188.

Manning, D., Ethell, S., Donovan, T. and Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography*, **12**, 134–142.

Martin, T., Burgman, M., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. and Mengersen, K. (2012). Eliciting expert knowledge in conservation science. *Conservation*, **26**, 29-38.

Mayer, F. S. and Frantz, C.M. (2004). The connectedness to nature scale: a measure of individuals' feeling in community with nature. *Journal of Environmental Psychology*, **24**, 503-515.

Megreya, A.M. and Bindemann, M. (2017). A visual processing advantage for young-adolescent deaf observers: Evidence from face and object matching tasks. *Scientific Reports*, **7**, 41133.

National Biodiversity Network [Online] (2011). *Improving Wildlife Data Quality*. NBN Trust, Nottingham. Available from: https://nbn.org.uk/wp-content/uploads/2016/02/NBN-Imp-Wildlife-Data-Quality-web.pdf [Accessed 1 March 2018].

Phillips, P.J. and O'Toole, A.J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing* **32**, 74-85.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, **30**, 685-696.

Powney, G.D. and Isaac, N.J.B. (2015). Beyond maps: a review of the applications of biological records. *Biological Journal of the Linnean Society,* **115**, 532–542.

Robertson, D.J., Noyse, E., Dowsett, A.J., Jenkins, R. and Burton, A.M. (2016). Face recognition by Metropolitan Police super-recognisers. *PLoS One* **11**, e0150036.

Rubin, J. and Chisnell, D. (2008). *Handbook of Usability Testing* 2nd ed. Wiley Publishing Inc., Indianapolis.

Russell, R., Duchaine, B. and Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, **16**, 252–257.

Ruiz-Gutierrez, V., Hooten, M.B. and Campbell Grant, E.H. (2016). Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, **7**, 900-909.

Scorolli, C. and Borghi, A. (2015). Square bananas, blue horses: the relative weight of shape and color in concept recognition and representation. *Frontiers in psychology*, **6**.

Shanks, D.R. (2005). Implicit learning. In: Lamberts, K. and Goldstone, R. eds. *Handbook of Cognition*. Sage Publications, London. pp.202-220.

Shipman, A.C.S. and Boster, J.S. (2008). Recall, similarity judgment, and identification of trees: A comparison of experts and novices. *Ethos*, **36**, 171–193.

Silvertown, J., Harvey, M., Greenwood, R., Dodd, M., Rosewell, J., Rebelo, T., Ansine, J. and McConway, K. 2015. Crowdsourcing the identification of organisms: a case-study of iSpot. *ZooKeys*, **480**, 125–146.

Solow, A., Smith, W., Burgman, M., Rout, T., Wintle, B. and Roberts, D. (2011). Uncertain sightings and the extinction of the ivory-billed woodpecker. *Conservation Biology*, **26**, 180–184.

Strassheim, H. (2015). Politics and policy expertise: towards a political epistemology. In: Fischer, F., Torgerson, D., Orsini, M. and Durnova, A. eds. *Handbook of Critical Policy Studies*, Cheltenham. pp. 319-340.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. and Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data* **2**:150026.

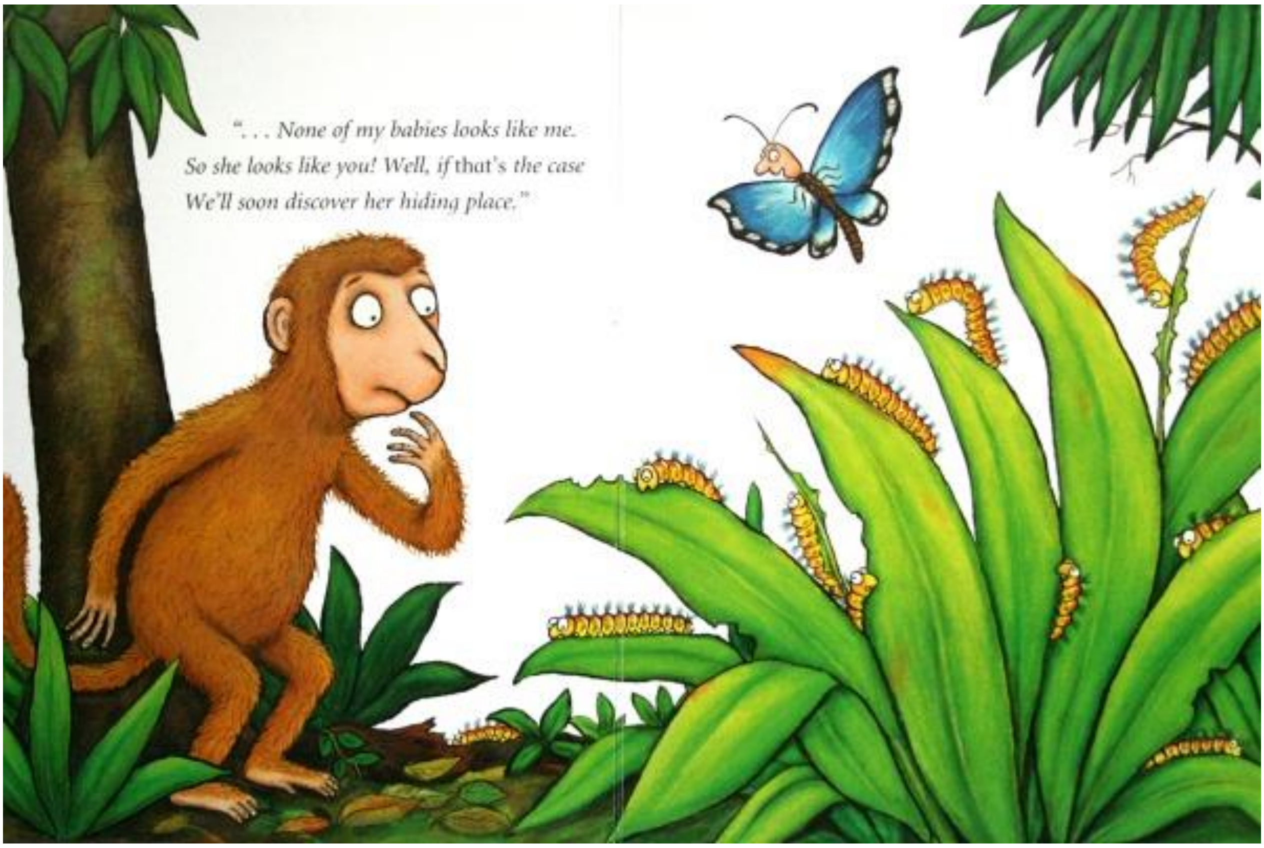
Taubert, J., Weldon, K.B. and Parr, L.A. (2016). Robust representations of individual faces in chimpanzees (*Pan troglodytes*) but not monkeys (*Macaca mulatta*). *Animal Cognition*, **20**, 1–9.

Termeer, C., Dewulf, A. and Breeman, G. (2013). Governance of Wicked Climate Adaptation Problems. In: Knieling, J., Leal Filho, W. eds. *Climate change governance*. Springer, pp. 27–39.

Verma, A., van der Wal, R. and Fischer, A. (2016). Imagining wildlife: new technologies and animal censuses, maps and museums. *Geoforum* ,**75**, 75–86.

White, D., Kemp, R.I., Jenkins, R., Matheson, M. and Burton, A.M. (2014). Passport officers' errors in face matching. *PloS One*, **9**, e103510.

Wilkinson, C. and Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science & Justice*, **49**, 191–196.



Monkey Puzzle © Julia Donaldson and Axel Scheffler 2000
Macmillan Publishers