Title:

Accuracy of identifications of mammal species from camera trap images: a northern Australian case study

Author details:

Larissa C. Potter

Corresponding author; larissa.potter94@gmail.com; (mobile) 0477 808 361; Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT, 0909, Australia

Christopher J. Brady

Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT, 0909, Australia

Brett P. Murphy

NESP Threatened Species Recovery Hub, Charles Darwin University, Darwin, NT, 0909, Australia

Running title: Accurate mammal identifications from camera traps

Acknowledgements

Research was supported by Tiwi Plantation Corporation, Plantation Management Partners, The Tiwi Land Council and The Australian Research Council (DE130100434). We would like to thank Ian Radford, Anna Miller, Rebecca Diete, William Ross and Hugh Davies for providing camera trap images. We are grateful to everyone who participated in our survey. This study was conducted with Human Ethics approval from Charles Darwin University (H17001).

MISS LARISSA CARMEL POTTER (Orcid ID : 0000-0002-6244-9941)

Article type : Original Article

Accuracy of identifications of mammal species from camera trap images: a northern Australian case study

Larissa C. Potter[1,2]
Christopher J. Brady[2]
Brett P. Murphy[2,3]

Abstract

Camera traps are a powerful and increasingly popular tool for mammal research, but like all survey methods, they have limitations. Identifying animal species from images is a critical component of camera trap studies, yet while researchers recognise constraints with experimental design or camera technology, image misidentification is still not well understood. We evaluated the effects of a species' attributes (body mass and distinctiveness) and individual observer variables (experience and confidence) on the accuracy of mammal identifications from camera trap images. We conducted an internet-based survey containing 20 questions about observer experience and 60 camera trap images to identify. Images were sourced from surveys in northern Australia and included 25 species, ranging in body mass from the Delicate mouse (*Pseudomys delicatulus*, 10 g) to the Agile wallaby (*Macropus agilis*, >10 kg). There was a weak relationship between accuracy of mammal identifications and observer experience. Accuracy was highest (100%) for distinctive species (e.g. Short-beaked echidna [*Tachyglossus aculeatus*]) and lowest (36%) for superficially non-distinctive mammals (e.g. rodents like the Pale field-rat [*Rattus tunneyi*]). There was a positive relationship between accuracy of identifications and body mass. Participant confidence was highest for large and distinctive mammals, but was not related to participant experience level. Identifications made with greater confidence were more likely to be accurate. Unreliability in identifications of mammal species is a significant limitation to camera trap studies, particularly where small mammals are the focus, or where similar-looking species co-occur. Integration of camera traps with conventional survey techniques (e.g. live-trapping), use of a reference library or computer-automated programs are likely to aid positive identifications, while employing a confidence rating system and/or multiple observers may lead to collection of more robust data. Although our study focussed on Australian species, our findings apply to camera trap studies globally.

Key words: camera trap, northern Australia, species identification, small mammal, wildlife survey

Introduction

Over the last three decades, the number of camera trap studies for detecting mammals has risen dramatically (Meek *et al.* 2015). This is partly a response to increased availability and affordability of commercial devices (Meek *et al.* 2015; Tobler *et al.* 2008), but also a result of advantages of camera traps over other sampling methods (De Bondi *et al.* 2010). However, like all survey methods, camera traps have inherent limitations, and it is crucial they are understood and acknowledged (Claridge and Paull 2014; Meek *et al.* 2015b; Meek *et al.* 2014). Currently, our understanding of the constraints of camera trapping is limited, particularly how these constraints affect our capacity to obtain unbiased and ecologically meaningful data (Burton *et al.* 2015; Meek *et al.* 2015b; Newey *et al.* 2015). This is particularly important given the increasing use of camera traps to aid management and conservation decisions (Burns *et al.* 2018; Comer *et al.* 2018; Jenks *et al.* 2011).

In a review of the Australian camera trap literature between 1991 and 2003, Meek *et al.* (2015b) found few studies acknowledging or discussing impacts of camera trap limitations on the outcome of results. When mentioned, researchers tend to focus on limitations of cameras themselves, including false triggers, battery life and reliability (Glen *et al.* 2013; Moseby and Read 2014), or experimental design elements like camera orientation (De Bondi *et al.* 2010; Smith and Coulson 2012), detection power (Nelson *et al.* 2014) and comparisons to other sampling methods (Ballard *et al.* 2014; Swan *et al.* 2014b). However, Meek *et al.* (2015b) classified the pitfalls of camera trapping into three broad categories: cameras, animals and observers. While the limitations of camera technology are increasingly recognised, the effect of species' attributes and human ability are still not well understood. Although camera technology is automated, the identification of images is generally done manually, and is strongly influenced by human ability (Ballard *et al.* 2014; Burns *et al.* 2018; Vernes *et al.* 2014). For example, in their examination of misidentification of small rodents in Victoria, Burns *et al.* (2018) found that accuracy in species identifications was species-specific and conditional on image type (white-flash vs. infrared), but the relationship between accuracy and experience was complicated, with the conclusion that species identification appears to be an innate skill.

Species identification from camera trap images is potentially difficult, introduces inherent error and may be biased by observer skill and experience (Dundas *et al.* 2014; Meek *et al.* 2013). Difficulty with species identification may also be affected by the presence of superficially similar, sympatric species (Claridge *et al.* 2010; Meek *et al.* 2013; Oliveira-Santos *et al.* 2010). For example, Meek and Vernes (2016) remarked on the difficulty in discriminating between eight sympatric rodents from the family Muridae, while Claridge *et al.* (2010) reported difficulty distinguishing small and superficially similar marsupials from the genus *Antechinus* (family Dasyuridae).

Due to the growing importance of camera trap survey data for conservation and management, it is imperative to understand factors which may affect accuracy of mammal identifications. The aim of this study was to investigate the effect of two species-level attributes, animal body mass and superficial distinctiveness (i.e. size, body shape, pelage colouration or patterning), and two observer-level attributes, experience and confidence, on accuracy of identification of mammal species from camera trap images. We predicted that accuracy and observer confidence in identifications would be lowest for small, non-distinctive species, in line with personal experience and the literature (Claridge *et al.* 2010; Meek and Vernes 2016). Additionally, we expected that more experienced observers would demonstrate higher accuracy and confidence levels.

Materials and methods

Collection of camera trap images

Sixty camera trap images of 25 native terrestrial mammal species (Table 1) were collated from six research projects across northern Australia, including coastal regions of the Northern Territory (NT) and the Kimberley region, Western Australia (WA). Individuals were identified to species level by researchers involved in each project (Corey *et al.* 2013; Davies *et al.* 2017; Diete *et al.* 2017) using image sequences, local knowledge, and confirmation from trap records.

Survey design

To assess accuracy of mammal identifications by wildlife scientists and enthusiasts, an internet-based survey was developed using the website SurveyMonkey™ (www.surveymonkey.com). Respondents were canvassed through Twitter (www.twitter.com), LinkedIn (www.linkedin.com; Australian Ecologists and Environmental Professionals page) and the Facebook (www.facebook.com) groups: Australian Mammal Society, Australian Mammal Identification and Wildlife Camera Trapping. Additionally, colleagues and professional ecologists were emailed directly and asked to distribute the survey through their networks. Due to the public nature of social media, a range of experience levels were obtained. The survey was open between 20 December 2016 and 3 March 2017.

The survey contained 20 questions regarding observer experience, followed by 60 camera trap images to be identified. Experience questions were divided into three sections: live trapping experience (n=8), camera trapping experience (n=6) and camera trap image identification experience (n=6). For image identification, respondents were asked to identify individuals to species, and assign a confidence rating to their identification. Confidence rating was a dropdown menu containing the following categories: >95% (“definite”); 86-94% (“pretty sure”); 66-85% (“probable”); 50-65% (“possible”); 36-49% (“not sure”); and <35% (“no idea”). While a numeric answer was preferable for analysis, words were used in combination to provide respondents with a better indication of what was meant by the confidence rating.

To mimic general wildlife surveys, a range of image types were used, including day and night, colour (white flash) and monochrome (infrared). All images were non-blurry, from horizontally placed camera traps and contained a single species with >90% of an individual visible within the field of view. Although it would have been preferable to have the full range of image types for each species, this was not possible. Additionally, location descriptions were provided with each image as a practitioner would typically have access to this information to assist in differentiating similar species. No single image was repeated, but multiple images of most species were included (Table 1) to reduce the likelihood that a species was misidentified due to low quality imagery. Respondents were asked to identify the first 24 images as a minimum because these contained one of each species (except Short-beaked echidna [*Tachyglossus aculeatus*] due to survey page design). The remaining 36 images were randomised so that if respondents did not complete the survey, the same images were not excluded each time.

Statistical Analysis

All analyses were conducted in the computer program R, version 3.3.2 (R Core Team, 2016). Prior to analysis, it was necessary to standardise comment-type responses and convert length of time answers to a single value. Due to the number of questions used to gauge observer experience (n=20), only the three broadest length of time responses were used: i.e. 'years trapping and handling mammals in Australia', 'years using camera traps' and 'years identifying wildlife in camera trap images'. Since these responses are not necessarily independent, covariance was examined with the Pearson Correlation Coefficient (Quinn and Keough 2002). All three responses were strongly correlated and thus combined into a single metric of 'experience'. For each respondent, we took the midpoint of 'years trapping' and the largest value of either 'years camera trapping' or 'years of image identification'. This approach to deriving an experience metric had the advantage over other methods (e.g. Principal Component Analysis) of providing a metric with interpretable units (i.e. years of experience).

To deal with variability in species identifications, responses were converted to binary variables with correct (1), incorrect (0) or non-response (blank) codes. Responses were classed as correct if the correct scientific or common name of the species was provided unambiguously, and incorrect for inappropriate species names, general terms (such as 'rodent') or invalid answers. Blank answers were assumed to be a non-attempt and excluded from the analysis.

Body mass for each species was taken from Van Dyck *et al.* (2013) as either the average value, or the midpoint of the male and female range provided (Table 1). Additionally, each species was assigned a 'distinctiveness' index, based on the number of species within its genus (Table 1) (from Van Dyck *et al.* (2013)). Distinctiveness index was calculated as a percentage: *distinctiveness* = ((23 – S)/23) x 100, where S is the total number of species in a particular genus and 23 the maximum number of species in the rodent genus *Pseudomys*. The index was rescaled using this maximum, so that larger values indicated greater distinctiveness.

Binomial generalised linear mixed models (GLMMs) with a logit-link function were developed using the *glmmML* package to examine predictors of accurate identifications. This modelling approach allowed respondent to be included as a random effect. Since predictor variables were measured in different units, and to allow interpretation of the effect size of each, body mass, experience and distinctiveness were centred and standardised, and rows with missing values omitted (Quinn and Keough 2002).

An information-theoretic approach (Burnham and Anderson 2003) was used to compare a set of candidate models developed for each response variable. Sixteen models were developed, representing all possible combinations of experience, body mass, distinctiveness and confidence.

Models within a set were ranked using the robust second-order form of Akaike's Information Criteria (AIC_c), and Δ_{AICc} (difference between AIC_c of a model and the minimum AIC_c in the candidate set) values calculated (Burnham and Anderson 1998). Additionally, Akaike weights (ω_i) were computed as a measure of the probability of a model being the best in the candidate set. Since AIC-based methods do not present information on the variance explained by a model, D^2, or the proportion of deviance explained by each model compared to the null model, was calculated (Nakagawa and Schielzeth 2013).

To examine variables influencing observer confidence, a set of candidate models were developed containing body mass, experience and distinctiveness. Confidence was treated as an ordinal response with 'no idea' < 'not sure' < 'possible' < 'probable' < 'pretty sure' < 'definite'. Since this is a multinomial response, models were run as proportional odds logistic regression (command *polr*) in the *MASS* package. Models were ranked using AIC_c, and Δ_{AICc}, ω_i and D^2 were calculated and used for model evaluation.

Results

A total 178 respondents answered the experience section and 129 attempted image identification. Of the 129, 83% had trapped and handled mammals in Australia, with experience ranging from 0 to >40 years. However, only 40% had done so in northern Australia. Similarly, 82% of respondents had used camera traps, with 37% deploying them in the study region and 89% had identified mammals from camera trap images. The most experienced respondents had used camera traps for 20 years and spent up to 14 years identifying mammals from their images.

Accuracy of mammal identifications

Accuracy of species identifications was highest for larger mammals, while smaller species, like the rodents, were often misidentified (Table 1). A positive relationship was found between accuracy of responses and a species' body mass (Figure 1), with accuracy increasing from 65% for the smallest mammals (<30 g) to 90% for the largest species (>10 kg) (D^2 = 0.16; Figure 1).

A positive relationship was observed between species distinctiveness and accuracy of identifications (Figure 2). A non-distinctive species had a lower predicted accuracy (60%), compared to a greater proportion (75%) of correct responses for a more distinctive species (Figure 2).

There was no distinct relationship between observer experience and accuracy of mammal identifications (Figure 3). However, the model predictions demonstrate that observers with no experience had an accuracy of 68%, while respondents with the greatest experience (24 years) had an accuracy of 80% (Figure 3).

The above trends in accuracy were supported by the modelling approach. Body mass, distinctiveness, experience and confidence were important factors to accurate identifications. This model explained only 16% of the data, but was the best model in the candidate set (ω_i = 1.00) (Table 2).

Confidence

A strong positive relationship was modelled between confidence (as a predictor) and proportion of correct responses (Figure 4). An increase in confidence from 'no idea' (35%) to 'definite' (95%) corresponded to a predicted rise in accuracy from 22% to 83% (Figure 4). Model selection showed that body mass, experience and distinctiveness influenced the confidence rating of a respondent, with this model having a high probability of being the best in the candidate set (ω_i = 1.00) (Table 3). However, this model explained only 9% of the deviance. Model predictions demonstrated a strong positive relationship between body mass and confidence, with 25% of responses being 'definite' for small mammals (10 g), to 85% 'definite' for the largest mammals (>10 kg) (Figure 5a). Modelled confidence as a function of observer experience showed no obvious relationship, with the proportion of 'definite' responses only increasing slightly from 55% to 65% (Figure 5b). Additionally, distinctive animals had a higher probability of a 'definite' rating (75%) compared to a less distinctive species (45%) (Figure 5c).

Discussion

Understanding limitations associated with camera traps is essential for obtaining robust data (Burton *et al.* 2015; Meek *et al.* 2015b; Newey *et al.* 2015). Our findings demonstrate that uncertainty in identifying mammals to species level is a genuine limitation of camera trap studies. Correct identifications and corresponding confidence levels were significantly higher for larger, more distinctive species while experience was not a strong predictor of accuracy or confidence. However, respondents who were more confident were more likely to be correct .

Camera traps are increasingly employed as the sole survey method for small to medium-sized mammals (<5 kg body mass) (Meek and Vernes 2016). However, our results demonstrate that practitioners' capacity to accurately identify such fauna from camera trap images is limited, especially for non-distinctive species. Accuracy for rodents, such as the Pale field-rat (*Rattus tunneyi*), were below 40% (Table 1) and the small dasyurids, Red-cheeked dunnart (*Sminthopsis virginiae*) and Butler's dunnart (*Sminthopsis butleri*), were often confused, 51% and 55% accuracy respectively (Table 1). In comparison, the Dingo (*Canis dingo*) and Short-beaked echidna, both large and distinctive species, were always correctly identified (100%) (Table 1). These results support the observations of Meek and Vernes (2016) and Claridge *et al.* (2010), who reported that distinguishing small rodent and dasyurid species was problematic.

While our index (based on the number of species in a genus) provided an objective proxy for distinctiveness, another approach would be to characterise distinctiveness based on the presence of conspicuous morphological features, such as spots (e.g. Northern quoll [*Dasyurus hallucatus*]) or an obvious white tail tip (e.g. Black-footed tree-rat [*Mesembriomys gouldi*]). Where obvious features were lacking within a genus (e.g. the Golden bandicoot [*Isoodon auratus*] compared to the sympatric Northern Brown bandicoot [*Isoodon macrourus*]), misidentification occurred (38%) (Table 1). Previous studies have also reported low accuracy in the identification of sympatric species of bandicoots from camera trap images (Claridge *et al.* 2010; Meek *et al.* 2013). The study by Meek *et al.* (2013), is one of the few to investigate the complexities of species identifications from camera trap images and found overall accuracy of small and medium-sized mammal identification to be relatively low (44.5%). In comparison to our study, however, Meek *et al.* (2013) included fewer species, only 30 experts and did not examine experience or confidence levels. Similarity between genera (e.g. the rodents *Psuedomys*, *Melomys* and *Rattus*) and distinctiveness within a genus (e.g. *Macropus*), were not captured by our distinctiveness index. Other approaches, such as an internet poll with camera trap practitioners, or a rating based on personal perspective, may have been more appropriate but are subjective and have their own limitations.

Difficulty distinguishing small- to medium-sized mammals is likely a result of both morphological and behavioural factors. Diagnostic features such as head-body to tail ratio, pelage colour and body shape are often used to distinguish species (Burns *et al.* 2018;

Claridge and Paull 2014; De Bondi *et al.* 2010). For example, when investigating whether the Hastings River mouse (*Pseudomys oralis*) could be differentiated from sympatric small mammals, Meek and Vernes (2016) used a key facial feature, the ‘Roman’-shaped nose, for identification. Similarly, Burns *et al.* (2018) demonstrated pelage colouration and morphology were important for distinguishing the smoky mouse (*Pseudomys fumeus*) and New Holland mouse (*Pseudomys novaehollandiae*) from sympatric rodents. However, visibility of such features is highly dependent on image quality and animal size (Burns *et al.* 2018). Lighting, camera-to-target distance and animal position, are factors which can mask distinguishing features (Meek *et al.* 2013; Oliveira-Santos *et al.* 2010). Our selected images included a range of lighting conditions - diurnal, nocturnal, white-flash and infrared (Supplementary Table S3). Due to the small number of images and the fact not all conditions were available for each species, we were not able to account for this variable in our models. This is an important limitation of our study as Burns *et al.* (2018) recently found that the effect of image type on accuracy of identifications can be significant. In their investigation, the authors found that white-flash (and hence colour) was crucial for identifying *P. fumeus*, while observers were more accurate identifying *P. novaehollandiae* from infrared images (where morphology was more distinctive). Additionally, small- to medium-sized mammals generally move faster through camera trap detection zones (Glen *et al.* 2013; Swan *et al.* 2014b), reducing the number of images, and the likelihood of clear images being obtained. For this survey, we selected only single, high-quality images of each species, but image sequences, rather than a single image, may allow several distinctive features and movement patterns to be observed (Claridge and Paull 2014; Meek *et al.* 2013), thus aiding with accurate identifications.

While some studies mention difficulty identifying small- to medium-sized mammals from camera trap images, few discuss the implications this may have on results (Meek *et al.* 2014). For example, Urlus *et al.* (2014) comment on monochrome images being harder for distinguishing small- to medium-sized mammals, but do not discuss how this may have affected the detectability of five mammal species examined. Similarly, Vernes *et al.* (2014) acknowledged that mammal species were “identified where possible”, but that this was sometimes impossible when individuals were too small, particularly shrews of the genus *Sorex*. Despite including ‘unknown small mammal’, ‘unknown large mammal’ and ‘unknown animal' in their results section, image identification was not discussed. This highlights that while species identification may not always be an issue, where it is problematic, it requires consideration.

Our results show that experience was not a strong predictor of accurate mammal identifications from camera trap images . This was unexpected because in many studies, images are sent to experts for verification (Falzon *et al.* 2014; Tobler *et al.* 2008). For example, while inventorying ground-dwelling mammals in southern Australia, Antos and Yuen (2014) captured an image of a rodent resembling a Broad-toothed rat (*Mastacomys fuscus*). They reported that the image was "awaiting confirmation from experts", and follow-up live-trapping was to be carried out. Although we hypothesized that experience would predict accuracy, the contrasting results are understandable. Despite expertise, distinguishing some species can be difficult even when in the hand (Falzon *et al.* 2014; Meek and Vernes 2016). While most respondents had prior experience with Australian mammals, including trapping, camera trapping and image identification (83%, 82% and 89% respectively), fewer respondents had trapped (40%), or employed cameras (37%), in northern Australia. Thus, respondents with a high level of experience may not have encountered the species included in our survey. This may have influenced accurate identifications, as prior experience with local species is likely to improve accuracy of identifications. While indication of morphological characteristics can be obtained from a field guide, seeing an animal up-close is a distinct advantage, because variability between individuals of a species may be high. Furthermore, camera trap practitioners generally work with large numbers of images, often seeing target species repeatedly. Since we only included a few images (in some cases only a single image) of a species, this may be a contributing factor to low accuracy. A greater number of images could have been included, however we felt that the length of the survey would have reduced the number of respondents.

Type of experience (e.g. consultant or naturalist), or how recently a respondent had handled or used camera traps, may have affected accuracy of identifications. However, these measures of experience were not examined in relation to accuracy for this study. This is partly because respondents could select multiple answers to the 'type of experience' questions, but also because time is more likely to be a better predictor of experience. In this modern era, information on mammal species, including images and descriptions, are widely accessible to most members of the public. Thus, a dedicated respondent with access to such resources may accurately identify fauna from images regardless of their experience with mammal identification. This may have important implications for camera trap projects relying on volunteers for image identification. However, intimate knowledge of target species or study location is likely to be crucial for accurate identifications.

The strong positive relationship predicted between confidence and accuracy of identifications, demonstrates an important 'safeguard' to this limitation of camera trapping. Respondents with low accuracy were more likely to have a low confidence rating with their identification, regardless of their experience. This suggests that respondents recognise when they have a high likelihood of being incorrect. This is supported by the low confidence ratings for small, less-distinctive mammals, for which accuracy levels were low. We use the term 'safeguard', because recognizing when a species cannot be identified is more likely to reduce potential negative consequences of misidentification. For example, if an individual resembling a threatened species is captured in a low-quality image, there are two potential biases: either an observer could misidentify the species thinking that it is too rare to be considered, or identify it as the threatened species because a false-positive may be perceived as preferable to a false-negative. The consequences of this can be significant (Burns *et al.* 2018), as numerous mammals, especially in northern Australia, are considered threatened. For some of these species, such as the Northern hopping mouse (*Notomys aquilo*), camera traps are the most suitable survey method (Diete *et al.* 2016). Thus, the ability to accurately identify threatened mammals from camera trap images is critical for monitoring and management. Employing confidence ratings with species identifications in future camera trap studies is likely to improve robustness of data obtained. Confidence ratings may assist with determining images that require closer inspection, cautious interpretation, or a live-trapping program for confirmation. Furthermore, low confidence images may be excluded from analysis due to the potential for false positives/negatives (Meek *et al.* 2014). Employing multiple observers may also improve reliability of species identification (Oliveira-Santos *et al.* 2010), as the degree of agreement between observers may perform better than confidence as a measure of uncertainty for a given identification.

Indeed, in our survey, the majority answer for each image was in perfect agreement with the 'true' identification provided by the donor of the image.

Currently, species identification from camera trap images tends to be a *post hoc* process, whereby species are recorded as they appear in images, and when difficult-to-distinguish individuals arise, they may be sent to experts for verification (Antos and Yuen 2014; Tobler *et al.* 2008). While this may work in some cases (large or distinctive species), we suggest the adoption of an *a priori* approach. When practitioners are selecting experimental design, they should also determine a species list for the study location, particularly small or morphologically-similar species. An effort should be made to obtain images of these species prior to camera deployment, therefore creating a reference library; familiarisation with these images may aid identification. In our survey, only 31% of respondents used an image reference library, compared to 73% relying on field guides. Meek *et al.* (2013) found 57% used a reference library and 73% used field guides. Additionally, in some cases, camera traps may not be the most suitable survey approach and this needs to be determined prior to sampling (Meek and Vernes 2016).

Advances in ecology are not only assisted by novel concepts, robust experiments or understanding of environmental systems, but also with the development of technology (Burton *et al.* 2015). According to Young *et al.* (2018), however, advances in technology used for camera trap management, the process from image collection to data organised for analysis, is developing slowly. As such, image identification is still a mostly manual process (Burns *et al.* 2018; Norouzzadeh *et al.* 2018; Young *et al.* 2018; Yu *et al.* 2013). Automatic subject detection (determination of whether an animal is present) and automatic species recognition are still in their infancy. Yu *et al.* (2013) employed techniques from computer vision science to successfully (82% accuracy) identify 18 species from 7000 camera trap images, and Norouzzadeh *et al.* (2018) used deep neural networks from artificial intelligence to identify species with >93.8% accuracy. However, these approaches are not without limitations. Large image databases and correctly identified images are required to teach the program, disadvantageous for rare species and small datasets (Young *et al.* 2018). Since small and morphologically similar species are most difficult for human observers to identify, future automated identification software should focus on these species, employing a combination of pelage colouration, morphological and behavioural features (e.g. gait) (Burns *et al.* 2018; Yu *et al.* 2013).

Identification of fauna to species level is a crucial aspect of camera trap studies, and consequences of misidentification are potentially significant. Knowledge of species' distributions and behaviour are fundamental to management decisions, but are often hindered because many terrestrial mammals are cryptic, nocturnal or rare (Swan *et al.* 2014b). Therefore, camera traps are emerging as a crucial tool for surveying mammals. However, we found that accurate species identification is a significant limitation of this survey tool, particularly for studies which focus on small mammals, or superficially non-distinctive species. Development of computer-assisted programs and combining camera trapping with other survey methods (e.g. live-trapping), is likely to greatly improve accuracy of species identifications (Dundas *et al.* 2014; Norouzzadeh *et al.* 2018; Young *et al.* 2018; Yu *et al.* 2013). Although only northern Australian species were included in our survey, the results are likely to be applicable in any region with diverse small- or morphologically-similar mammal communities (Meek *et al.* 2013; Vernes *et al.* 2014).

Acknowledgements

Research was supported by Tiwi Plantation Corporation, Plantation Management Partners, The Tiwi Land Council and The Australian Research Council (DE130100434). We would like to thank Ian Radford, Anna Miller, Rebecca Diete, William Ross and Hugh Davies for providing camera trap images. We are grateful to everyone who participated in our survey. This study was conducted with Human Ethics approval from Charles Darwin University (H17001).

References

Antos M. J. & Yuen K. (2014) Camera trap monitoring for inventory and management effectiveness in Victorian national parks: tailoring approaches to suit specific questions. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 13-26. CSIRO Publishing, VIC.

Ballard G., Meek P. D., Doak S., Fleming P. J. S. & Sparkes J. (2014) Camera traps, sand plots and known events: what do camera traps miss? In: *Camera trapping in wildlife management* *and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 189-202. CSIRO Publishing, VIC.

Burnham K. P. & Anderson D. R. (1998) *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag New York Inc., USA.

Burnham K. P. & Anderson D. R. (2003) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Burns P. A., Parrott M. L., Rowe K. C. & Phillips B. L. (2018) Identification of threatened rodent species using infrared and white-flash camera traps. *Australian Mammalogy* **40**, 188-97.

Burton A. C., Neilson E., Moreira D., Ladle A., Steenweg R., Fisher J. T., Bayne E. & Boutin S. (2015) Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology* **52**, 675-85.

Claridge A. W. & Paull D. J. (2014) How long is a piece of string? Camera trapping methodology is question dependent. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 205-14. CSIRO Publishing, VIC.

Claridge A. W., Paull D. J. & Barry S. C. (2010) Detection of medium-sized ground-dwelling mammals using infrared digital cameras: an alternative way forward? *Australian Mammalogy* **32**, 165-71.

Comer S., Speldewinde P., Tiller C., Clausen L., Pinder J., Cowen S. & Algar D. (2018) Evaluating the efficacy of a landscape scale feral cat control program using camera traps and occupancy models. *Scientific Reports* **8**.

Corey B., Radford I., Carnes K., Hatherley E. & Legge S. (2013) North-Kimberley landscape conservation initiative: 2010-12 performance report. Department of Parks and Wildlife, Kununurra, Western Australia.

Davies H. F., McCarthy M. A., Firth R. S. C., Woinarski J. C. Z., Gillespie G. R., Andersen A. N., Geyle H. M., Nicholson E. & Murphy B. P. (2017) Top-down control of species distributions: feral cats driving the regional extinction of a threatened rodent in northern Australia. *Diversity and Distributions* **23**, 272-83.

De Bondi N., White J. G., Stevens M. & Cooke R. (2010) A comparison of the effectiveness of camera trapping and live trapping for sampling terrestrial small-mammal communities. *Wildlife Research* **37**, 456-65.

Diete R. L., Meek P. D., Dickman C. R. & Leung L. K.-P. (2016) Ecology and conservation of the northern hopping-mouse (*Notomys aquilo*). *Australian Journal of Zoology* **64**, 21-32.

Diete R. L., Meek P. D., Dickman C. R., Lisle A. & Leung L. K.-P. (2017) Diel activity patterns of northern Australian small mammals: variation, fixity, and plasticity. *Journal of Mammalogy*, 1-10.

Dundas S. J., Adams P. J. & Fleming P. A. (2014) Can camera trap surveys provide reliable population estimates for nondescript species? In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 173-9. CSIRO, Vic.

Falzon G., Meek P. D. & Vernes K. (2014) Computer-assisted identification of small Australian mammals in camera trap imagery. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 299-306. CSIRO Publishing, VIC.

Glen A. S., Cockburn S., Nichols M., Ekanayake J. & Warburton B. (2013) Optimising camera traps for monitoring small mammals. *PLoS ONE* **8**, e67940.

Jenks K. E., Chanteap P., Damrongchainarong K., Cutter P., Cutter P., Redford T., Lynam A. J., Howard J. & Leimgruber P. (2011) Using relative abundance indices from camera-trapping to test wildlife conservation hypotheses – an example from Khao Yai National Park, Thailand. *Tropical Conservation Science* **4**, 113-31.

Meek P. D., Ballard G.-A. & Fleming P. J. S. (2015b) The pitfalls of wildlife camera trapping as a survey tool in Australia. *Australian Mammalogy* **37**, 13-22.

Meek P. D., Ballard G.-A., Vernes K. & Fleming P. J. S. (2015) The history of wildlife camera trapping as a survey tool in Australia. *Australian Mammalogy* **37**, 1-12.

Meek P. D., Fleming P. J. S., Ballard G., Banks P. B., Claridge A. W., McMahon S., Sanderson J. & Swann D. E. (2014) Putting contemporary camera trapping in focus. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 349-56. CSIRO Publishing, VIC.

Meek P. D. & Vernes K. (2016) Can camera trapping be used to accurately survey and monitor the Hastings River mouse (*Pseudomys oralis*)? *Australian Mammalogy* **38**, 44-51.

Meek P. D., Vernes K. & Falzon G. (2013) On the reliability of expert identification of small-medium sized mammals from camera trap photos. *Wildlife Biology in Practice* **9**, 1-19.

Moseby K. E. & Read J. L. (2014) Using camera traps to compare poison bait uptake by invasive predators and non-target species. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 131-9. CSIRO Publishing, VIC.

Nakagawa S. & Schielzeth H. (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133-42.

Nelson J. L., Scroggie M. P. & Belcher C. A. (2014) Developing a camera trap survey protocol to detect a rare marsupial carnivore, the spotted-tailed quoll (*Dasyurus maculatus*). In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 271-9. CSIRO Publishing, VIC.

Newey S., Davidson P., Nazir S., Fairhurst G., Verdicchio F., Irvine R. J. & van der Wal R. (2015) Limitations of recreational camera traps for wildlife management and conservation research: a practitioner's perspective. *Ambio*, 624-35.

Norouzzadeh M. S., Nguyen A., Kosmala M., Swanson A., Palmer M. S., Packer C. & Clune J. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *PNAS* **115**.

Oliveira-Santos L. G. R., Zucco C. A., Antunes P. C. & Crawshaw Jr P. G. (2010) Is it possible to individually identify mammals with no natural markings using camera-traps? A controlled case-study with lowland tapirs. *Mammalian Biology Zeitschrift fur Saugetierkunde* **75**, 375-8.

Quinn G. P. & Keough M. J. (2002) *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge, UK.

Smith J. K. & Coulson G. (2012) A comparison of vertical and horizontal camera trap orientations for detection of potoroos and bandicoots. *Australian Mammalogy* **34**, 196-201.

Swan M., Di Stefano J., Christie F., Steel E. & York A. (2014b) Detecting mammals in heterogeneous landscapes: implications for biodiversity monitoring and management. *Biodiversity Conservation* **23**, 343-55.

Tobler M. W., Carrillo-Percastegui S. E., Leite Pitman R., Mares R. & Powell G. (2008) An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. *Animal Conservation* **11**, 169-78.

Urlus J., McCutcheon C., Gilmore D. & McMahon J. (2014) The effect of camera trap type on the probability of detecting different size classes of Australian mammals. In: *Camera* *trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 111-21. CSIRO Publishing, Vic.

Van Dyck S., Gynther I. & Baker A. (2013) *Field companion to the mammals of Australia*. New Holland Publishers, London.

Vernes K., Smith M. & Jarman P. J. (2014) A novel camera-based approach to understanding the foraging behaviour of mycophagous mammals. In: *Camera trapping in wildlife management and research* (eds P. Meek, P. Fleming, G. Ballard, P. Banks, A. Claridge, J. Sanderson and D. Swann) pp. 215-24. CSIRO Publishing, VIC.

Young S., Rode-Margono J. & Amin R. (2018) Software to facilitate and streamline camera trap data management: a review. *Ecology and Evolution*, 1-11.

Yu X., Wang J., Kays R., Jansen P. A., Wang T. & Huang T. (2013) Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing* **2013**.

Tables

Table 1 The 25 native mammal species and the number of images (N) of each included in an internet-based survey to assess accuracy of mammal identifications from camera trap images. The distinctiveness index (D) was an index calculated as: *distinctiveness* = ((23 – *S*)/23) x 100, where *S* = number of species in the genus and 23 the maximum number of species in a single genus (*Pseudomys*).

Species		Body mass (g)	Distinctiveness index (D%)	N	Proportion (%) of correct responses
Agile wallaby	*Macropus agilis*	15000	39.1	2	83
Dingo	*Canis dingo*	14000	95.7	3	100
Short-beaked echidna	*Tachyglossus aculeatus*	4 500	95.7	3	100
Short-eared /Wilkins rock wallaby	*Petrogale brachyotis/wilkinsi*	4 050	30.4	4	64
Common brushtail possum	*Trichosurus vulpecula*	2 625	87.0	2	96
Rock ringtail possum	*Petropseudes dahli*	1 640	95.7	1	64
Northern brown bandicoot	*Isoodon macrourus*	1 600	87.0	3	69
Scaly-tailed possum	*Wyulda squamicaudata*	1 450	95.7	2	60
Monjon	*Petrogale burbidgei*	1 258	30.4	2	47
Black-footed tree-rat	*Mesembriomys gouldii*	716	91.3	3	79
Northern quoll	*Daysurus hallucatus*	597	82.6	3	75
Golden bandicoot	*Isoodon auratus*	485	87.0	1	38
Golden-backed tree-rat	*Mesembriomys macrurus*	267	91.3	3	69

Brush-tailed phascogale	*Phascogale tapoatafa*	193	87.0	2	71
Brush-tailed rabbit-rat	*Conilurus penicillatus*	153	91.3	3	62
Kimberley rock-rat	*Zyzomys woodwardi*	140	78.3	1	47
Sugar glider	*Petaurus breviceps*	127	82.6	2	89
Pale field-rat	*Rattus tunneyi*	86	43.5	3	36
Grassland melomys	*Melomys burtoni*	68	82.6	2	44
Northern hopping-mouse	*Notomys aquilo*	40	60.9	3	68
Common rock-rat	*Zyzomys argurus*	36	78.3	4	53
Red-cheeked dunnart	*Sminthopsis virginiae*	35	17.4	2	51
Butler's dunnart	*Sminthopsis butleri*	23	17.4	3	55
Delicate mouse	*Pseudomys delicatulus*	10	0	3	51
			Total	60	

Table 2 Candidate model selection results for factors affecting accuracy of mammal identifications from camera trap images in an internet-based survey. Respondent was included as a random factor.

ΔAICc is the difference between second-order Akaike Information Criterion of a model and the minimum AICc: ω_i is the Akaike weight, a measure of the probability of a model being the best in the candidate set: D^2 is the proportion of deviance explained by each model compared to the null.

Model	**AIC**	**Δ AIC*c***	ω_i	D^2
Body mass + confidence + experience + distinctiveness	3544.7	0.0	1.00	0.16
Body mass + confidence + distinctiveness	3570.4	25.7	0.00	0.15
Body mass + confidence + experience	3584.0	39.3	0.00	0.15
Body mass + confidence	3608.2	63.5	0.00	0.14
Confidence + experience + distinctiveness	3635.9	91.3	0.00	0.13
Confidence + distinctiveness	3657.2	112.5	0.00	0.13
Confidence + experience	3677.4	132.7	0.00	0.13
Confidence	3696.9	152.2	0.00	0.12
Body mass + distinctiveness + experience	3837.1	292.5	0.00	0.09
Body mass + distinctiveness	3868.5	323.8	0.00	0.08
Body mass + experience	3959.2	414.6	0.00	0.06
Body mass	3989.3	444.6	0.00	0.05
Distinctiveness + experience	4031.0	486.3	0.00	0.04
Distinctiveness	4056.6	511.9	0.00	0.04
Experience	4181.4	636.7	0.00	0.01
Null	4204.8	660.1	0.00	0.00

Table 3 Set of candidate models and model selection results to explain variation in observer confidence in species identifications from camera trap images.

ΔAICc is the difference between second-order Akaike Information Criterion of a model and the minimum AICc: ω_i is the Akaike weight, a measure of the probability of a model being the best in the candidate set: D^2 is the proportion of deviance explained by each model compared to the null.

Model	**AIC**	**Δ AIC_c**	ω_i	D^2
Body mass + distinctiveness + experience	9926.9	0.0	0.99	0.090

Body mass + distinctiveness	9960.2	33.3	<0.001	0.087
Body mass + experience	10059.0	132.1	<0.001	0.078
Body mass	10090.9	164.0	<0.001	0.075
Distinctiveness + experience	10445.4	518.5	<0.001	0.043
Distinctiveness	10467.3	540.3	<0.001	0.040
Experience	10887.1	960.2	<0.001	0.002
Null	10904.6	977.6	<0.001	0.000

Figure legends

Figure 1 Relationship between species body mass (g) and proportion of correct identifications from camera trap images in an internet-based survey. Body mass for each species was taken from Van Dyck *et al.* (2013) as either the average value, or the midpoint of the male and female range provided. The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. n=60 images (circles).

Figure 2 Relationship between distinctiveness of the species in a camera trap image and the proportion of correct responses in an internet-based survey. The distinctiveness index was calculated as: ((23–*S*)/23) x 100, where *S* is the number of species in a particular genus and 23 the maximum number of species in a single genus. The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. n=60 images (circles).

Figure 3 Relationship between observer experience (years) and proportion of correct species identifications in an internet-based survey. Experience was calculated as the mean of time trapping mammals and the largest value of camera experience (either years using camera traps or identifying camera trap images). The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. n=178 respondents (circles).

Figure 4 Predicted relationship between observer confidence and proportion of correct species identifications in an internet-based survey, where respondents were asked to assign a confidence rating to each identified image with the following categories: >95% ("definite"), 86-94% ("pretty sure"), 66-85% ("probable"), 50-65% ("possible"), 36-49% ("not sure") and <35% ("no idea"). Predictions were based on the model of best fit (thick line) and thin lines indicate 95% confidence intervals.

Figure 5 Modelled relationships between proportion of answers correct and observer confidence for a) species body mass (g), b) observer experience (years) and c) distinctiveness index ((23-*S*)/*23*) x 100, where *S* = number of species in genus and 23 = maximum number of species in a genus). Respondents were asked to assign a confidence rating to each identification in an internet-based survey with the categories: >95% ("definite"), 86-94% ("pretty sure"), 66-85% ("probable"), 50-65% ("possible"), 36-49% ("not sure") and <35% ("no idea"). Predictions were based on the best candidate model.

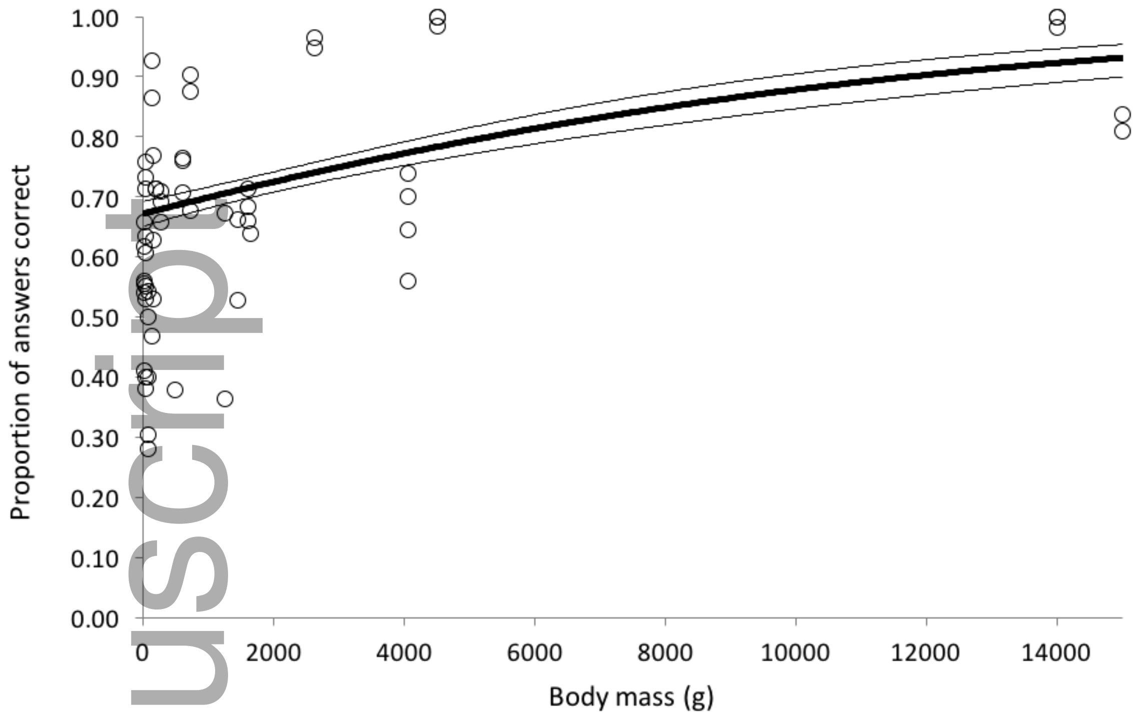

Figure 1 Relationship between species body mass (g) and proportion of correct identifications from camera trap images in an internet-based survey. Body mass for each species was taken from Van Dyck *et al.* (2013) as either the average value, or the midpoint of the male and female range provided. The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. *n*=60 images (circles).

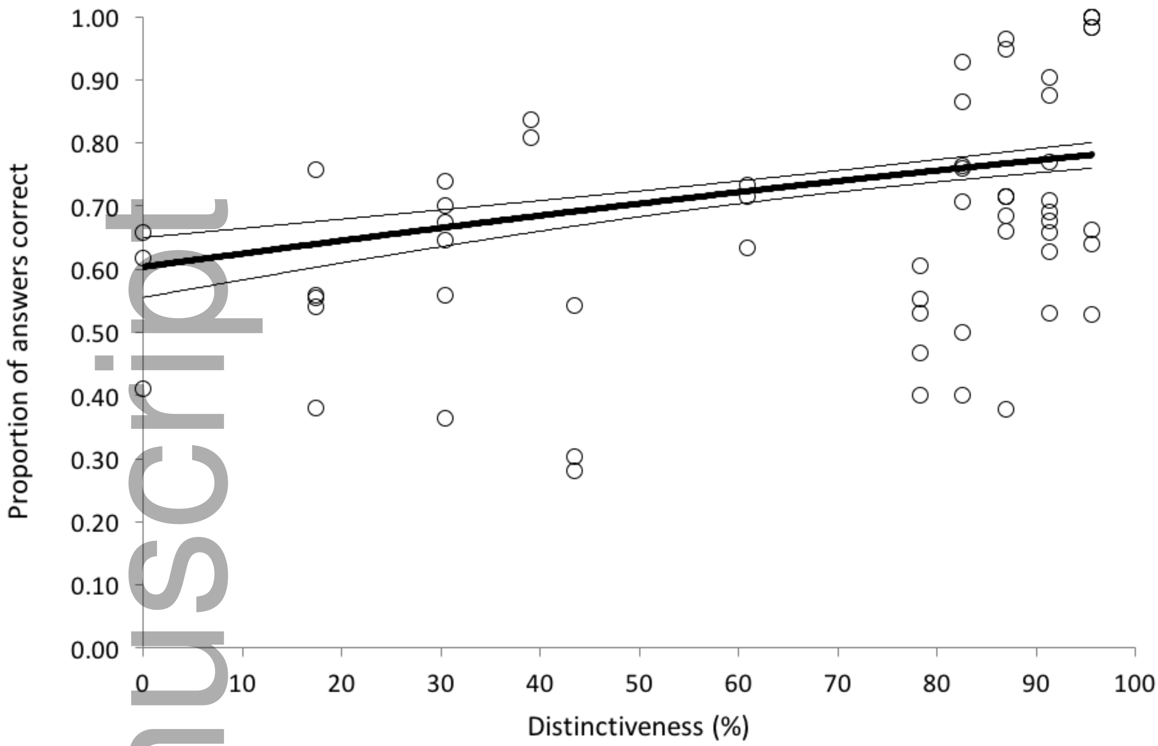

Figure 2 Relationship between distinctiveness of the species in a camera trap image and the proportion of correct responses in an internet-based survey. The distinctiveness index was calculated as: ((23–*S*)/23) x 100, where *S* is the number of species in a particular genus and 23 the maximum number of species in a single genus. The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. *n*=60 images (circles).

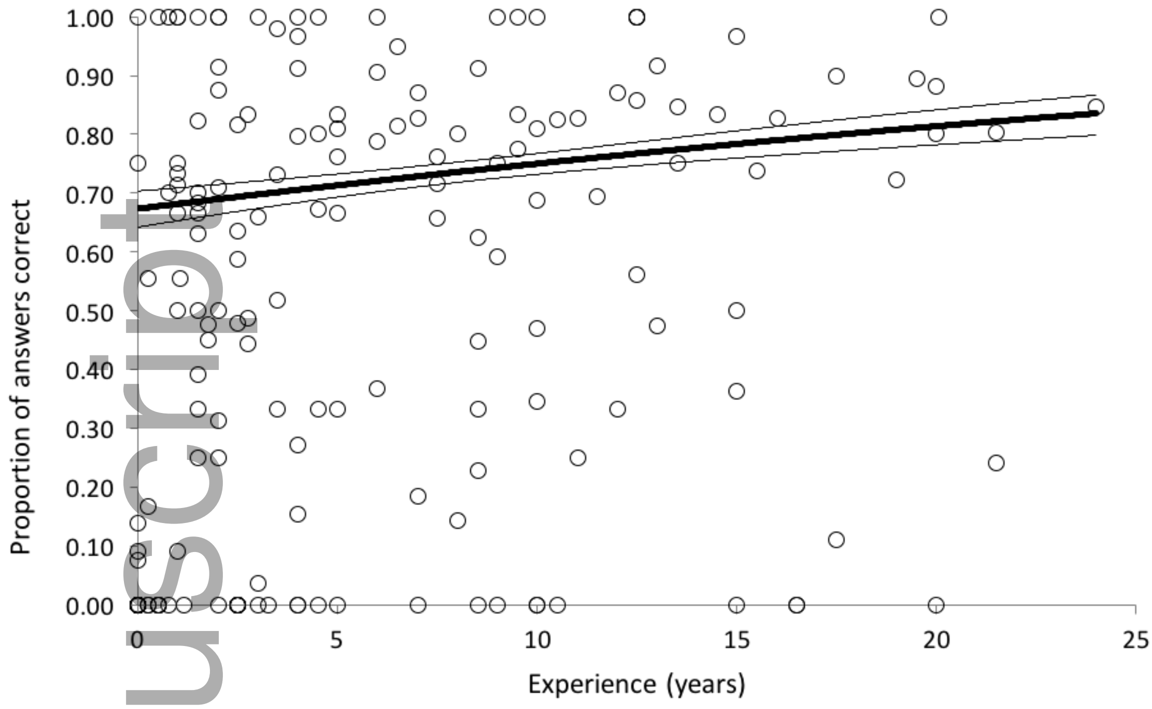

Figure 3 Relationship between observer experience (years) and proportion of correct species identifications in an internet-based survey. Experience was calculated as the mean of time trapping mammals and the largest value of camera experience (either years using camera traps or identifying camera trap images). The best model was used for predictions (thick line) and thin lines indicate 95% confidence intervals. n=178 respondents (circles).

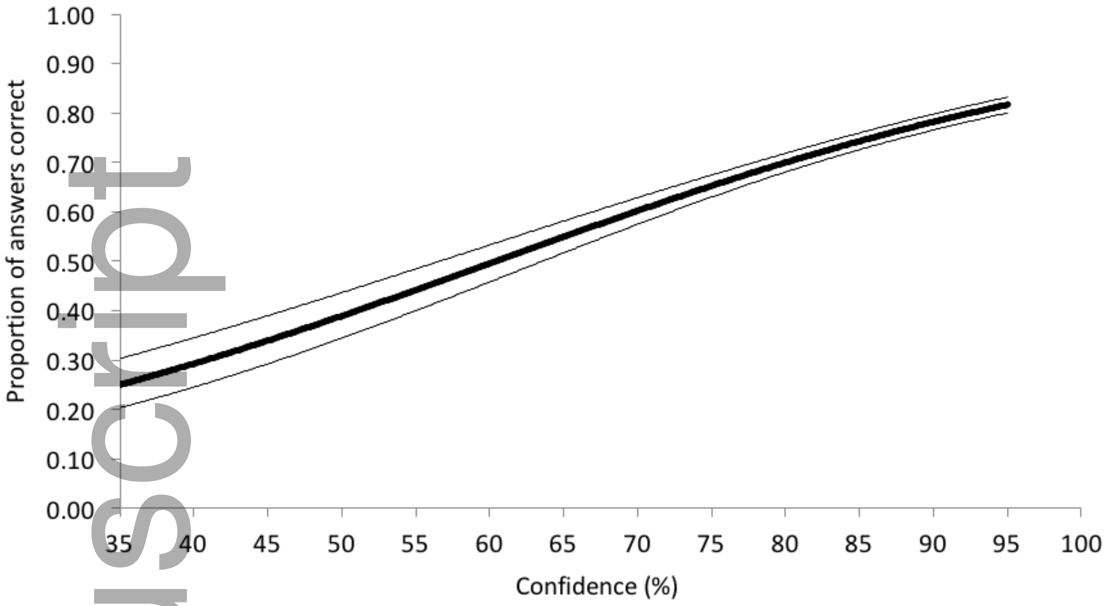

Figure 4 Predicted relationship between observer confidence and proportion of correct species identifications in an internet-based survey, where respondents were asked to assign a confidence rating to each identified image with the following categories: >95% ("definite"), 86-94% ("pretty sure"), 66-85% ("probable"), 50-65% ("possible"), 36-49% ("not sure") and <35% ("no idea"). Predictions were based on the model of best fit (thick line) and thin lines indicate 95% confidence intervals.

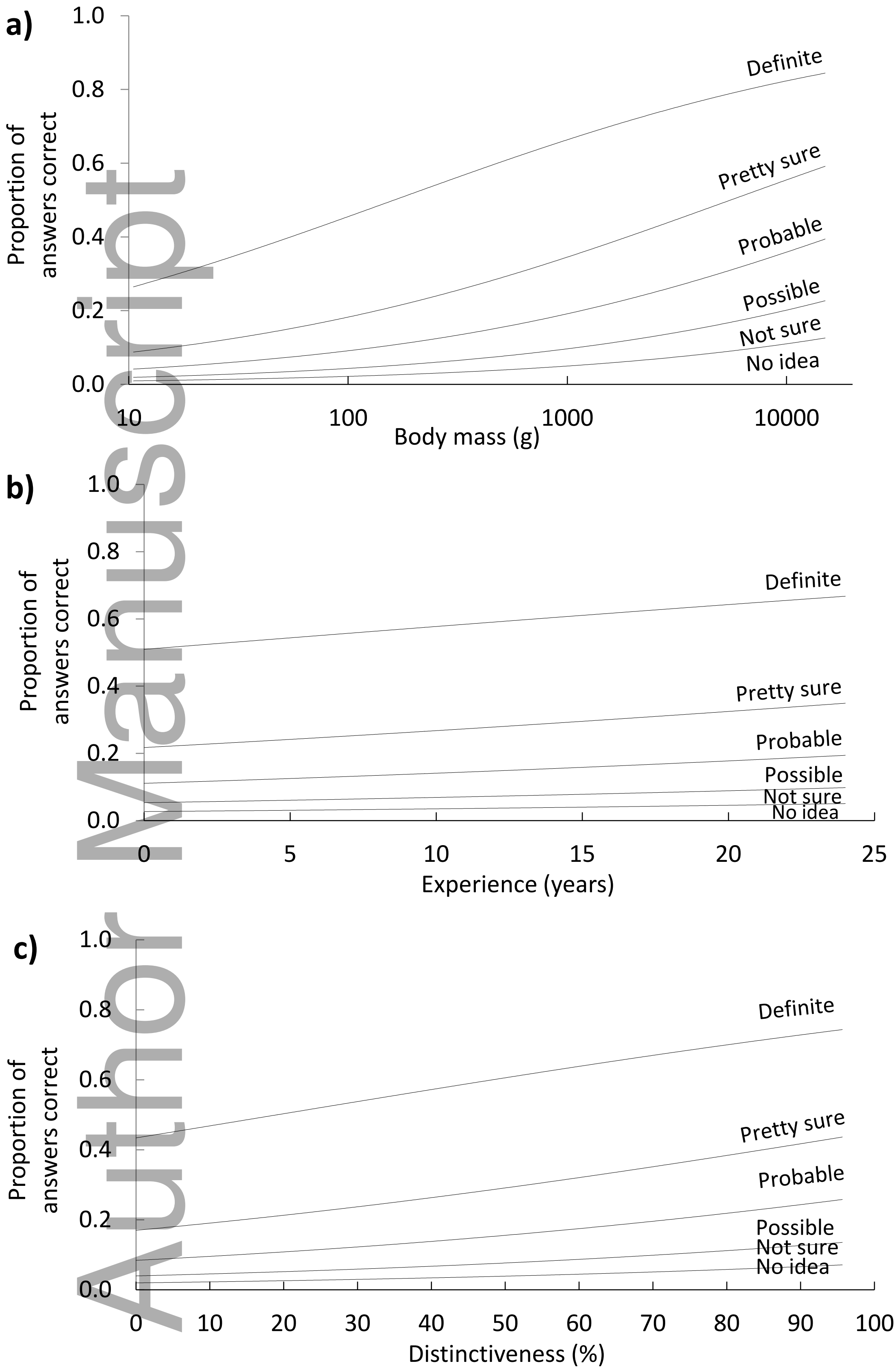

Figure 5 Modelled relationships between proportion of answers correct and observer confidence for a) species body mass (g), b) observer experience (years) and c)

distinctiveness index ($[(23-S)/23] \times 100$, where S = number of species in genus and 23 = maximum number of species in a genus). Respondents were asked to assign a confidence rating to each identification in an internet-based survey with the categories: >95% ("definite"), 86-94% ("pretty sure"), 66-85% ("probable"), 50-65% ("possible"), 36-49% ("not sure") and <35% ("no idea"). Predictions were based on the best candidate model.