Applied Animal Behaviour Science xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

Inconsistency in dairy calves' responses to tests of fearfulness

Rebecca K. Meagher*, Marina A.G. von Keyserlingk, Dax Atkinson, Daniel M. Weary

Animal Welfare Program, University of British Columbia, 2357 Main Mall, Vancouver, BC, V6T 1Z4, Canada

ARTICLE INFO

Article history:
Received 12 January 2016
Received in revised form 21 August 2016
Accepted 18 October 2016
Available online xxx

Keywords:
Fearfulness
Neophobia
Human-animal relationship
Well-being
Reliability
Validity

ABSTRACT

Fear is an important welfare problem for farm animals, including cattle. A variety of methods of assessing fear have been proposed, but the reliability and validity of these methods, and ways of improving these characteristics, have received little study. We conducted a series of experiments to assess the consistency of dairy calves' responses of novel objects and to humans, and to investigate factors that might improve reliability. In the first experiment, latency to touch a novel object had moderate reliability ($r_s = 0.54$), and latency to touch a stationary, familiar human had negligible reliability ($r_s = 0.26$). Experiment 2a used the same test protocols, but with a shorter interval between repeat testing and using different stimuli in the two novel object tests; this change did not improve reliability (e.g. $r_s = 0.29$ for the novel-object test). Reliability for this test was improved ($r_s = 0.58$) in Experiment 2b, when the same object was used in both tests rather than a truly novel object being used the second time. Experiment 2a found ceiling effects in the response to human test associated with the short period during which approach responses were recorded. High reliability was found in Experiment 2b, where the maximum test duration was doubled, but this effect not due to the extended duration. Experiment 3 assessed reliability of a response to human approach at the farm rather than individual level, in this case assessing responses to an unfamiliar person. The proportion of calves making contact with the person was not reliable ($r_s = 0.22$), but the proportion retreating from the person had moderate reliability ($r_s = 0.52$). Reliability was improved by excluding data from calves that had coughs on the day of testing. Conducting multiple tests per individual using different stimuli and reporting health status of the animals are recommended for future research and animal welfare assessment schemes that include measures of fear.

1. Introduction

Fear is widely recognized as a welfare concerns for cattle and other farm animals (e.g. Farm Animal Welfare Council, 2009; Hemsworth et al., 2000; Jones and Boissy, 2011). Fearful animals can also cause production and management challenges, including decreased productivity (e.g. Barnett et al., 1992; Hemsworth et al., 2000) and animals that are afraid of humans may be more dangerous to handle (Boivin et al., 1992; Hemsworth et al., 1989). Unfortunately, methods of assessing fear (a negative emotional state resulting from a perceived threat (Gray, 1987; Ennaceur, 2014)) and fearfulness (a personality trait characterized by a tendency to express fear when exposed to potentially threatening stimuli or situations) appear not to be well-validated and have uncertain reliability (Forkman et al., 2007). Of 112 papers published in this journal over a five-year period ending in August 2015 with fear* or anx* in the keywords, abstract or title; only 65 papers (or 58%) contained any form of the words reliable or repeatable anywhere in the text; and of these; only 15 actually estimated reliability. Measures also vary considerably across studies; making it difficult to extrapolate results from one approach to the next (Forkman et al., 2007).

The need for valid, reliable ways of assessing welfare in farm animals is widely recognized, to be used for example in assurance schemes for commercial farms (see Scott et al., 2001). Currently, fear is often assessed in farm animals through response to novelty (neophobia, although other factors such as exploratory motivation also influence the response), most commonly using a novel object test. Another common type of fear-related test is in response to humans (e.g. Forkman and Keeling, 2009), as fear of handlers may have a major impact on the lives of intensively farmed animals. Research published to date indicates that responses are not closely associated in these two contexts (e.g. Hegelund and Sorensen, 2007), and that separate measures may be needed. From the perspective of animal welfare, fearfulness and long-lasting states of fear are of special interest, meaning that we are especially interested in fear responses that are consistent over time. Unfortunately, test-retest reliability (also called repeatability) is often weak making it difficult to draw strong inferences from a single test.

* Corresponding author.
E-mail address: rkmeagher@gmail.com (R.K. Meagher).

http://dx.doi.org/10.1016/j.applanim.2016.10.007
0168-1591/

In cattle, for example, the novel object test was reported to be reliable within individuals between tests in at least two calf studies (using measures derived from factor analysis in Van Reenen et al., 2004; and approach latency in Bokkers et al., 2009), but was unreliable in older heifers and adult cows when tested using avoidance (Van Reenen et al., 2013), reactivity (Gibbons et al., 2009), number of interactions and time in proximity (Kilgour et al., 2006). Results have been mixed across a range of measures and ages in other studies (Graunke et al., 2013; MacKay et al., 2014). Even the methods of assessing 'repeatability' vary: while most studies replicate the test exactly using the same stimulus, others (e.g. Gibbons et al., 2009) instead assess consistency of response across different novel stimuli because there is no way to repeat a test and have it be truly novel (see e.g. Forkman et al., 2007 for a discussion of this problem). Nonetheless, the novel object test has face validity, meaning that it appears sensible based on our understanding of fear and comparisons with human behaviour, as judged by experts (e.g. Scott et al., 2001; Whay et al., 2003). It is also one of the few tests that has undergone some successful validation for cattle, suggesting it may be a true indicator of fear (based on correlation with other fear- and stress-related measures and pharmacological validation using anxiolytic drugs; e.g. Van Reenen et al., 2005, 2009). Confirming or finding ways to improve its reliability would thus be valuable.

Responses to humans (typically measured as approach or avoidance by the animal) are more consistently reported to be reliable (at the individual level in calves (Rousing et al., 2005) and cows (Gibbons et al., 2009; Turner et al., 2011)). However, some papers found moderate to high repeatability only for some measures and time periods (Haskell et al., 2012; Mazurek et al., 2011; Windschnurer et al., 2008; see also review of responses to humans by de Passillé and Rushen, 2005), and other studies have found no repeatability (Battini et al., 2011), although all of these studies depended on some measure of avoidance or retreat from a human. Fina et al. (2006) reported that reliability of responses to restraint differed depending upon the calves' initial responses, with calm individuals remaining calm across tests but fearful ones showing reduced fear over time.

Farm-level repeatability is also important for measures of approach or avoidance of humans, because this type of measure has been proposed for use in on-farm welfare assessments (e.g. Winckler et al., 2003, 2007), focussing on herd-level differences. Only a few papers have investigated farm-level repeatability of responses to humans, all in adult cows, and studies have sometimes confounded test-retest reliability with inter-observer reliability (e.g. Windschnurer et al., 2009a,b). In these tests (based upon avoidance of an approaching human) low to moderate reliability has been reported (De Rosa et al., 2003; Winckler et al., 2007). Reliability can also be estimated at the level of the pen or group (intermediate between individual and farm levels), and indeed some farm level estimates are based upon observations of a single pen. Only one study on calves has assessed the reliability of approach responses measured at the pen level, and this study reported high reliability (Bokkers et al., 2009; with similar results for an avoidance measure).

Even among papers that claim repeatability, correlations are sometimes low. For example, Turner et al. (2011) assessed repeatability across and within tests of fear of humans in beef cattle and found the proportion of variance explained by individual consistency ranged from 0.17 to 0.54. In fact, a meta-analysis of the personality literature in wild animals found an average repeatability (intraclass correlation coefficient) of only 0.37 (Bell et al., 2009), which is considerably below the level generally deemed acceptable (0.6 being a traditional standard in the human literature (e.g. Bruton et al., 2000; Mroczek 2007). In humans, typical correlations over long intervals (years) are often over 0.7 in adults (Mroczek, 2007). Conversely, correlation coefficients for children and college students were only 0.31 and 0.54 respectively, for major personality traits in one meta-analysis (Roberts and DelVecchio, 2000). It therefore seems likely that other juvenile animals, such as calves, may also show limited correlations in their fear responses over time.

The aims of the current study were to assess the individual-level test-retest reliability of versions of novel object and response to human tests, and the farm-level test-retest reliability of a response to human test. An additional aim was to identify factors that influence reliability, enabling refinements in protocols used in future research and on-farm welfare assessments. The factors investigated included consistency of the object used in the novel object test, test duration, and calf health. We also assessed inter- and intra-observer reliability (i.e. consistency between and within people recording the data) of the measures, as these are essential to obtaining test-retest reliability.

2. Materials and methods

2.1. Experiment 1

All of the research presented in this paper was approved by the University of British Columbia Animal Care Committee. In this experiment we used 32 Holstein bull calves, housed at the University of British Columbia Dairy Education and Research Centre. These calves also served in a concurrent study on the effects of early social housing, comparing individually housed calves (n = 10), pair-housed calves (n = 12), and calves kept in a complex social group with access to their dams (n = 10). More detail regarding these treatments is available in Meagher et al. (2015). Pens were cleaned once per week. Calves were offered 8 L of milk per day for the first 28 d, at which time the milk ration was reduced to 6 L over 3 d, always split between two daily feedings. This reduction was intended to stimulate solid feed intake. At approximately 58 d, calves were weaned over a 3-day period. Calves had ad libitum access to water throughout the experimental period, and access to grain (Hi-Pro Medicated Calf Starter) and a mixed ration beginning at day 5 ± 2. Health checks were performed weekly throughout the experimental period to assess symptoms of common illnesses, including respiratory and enteric disease. Calves were treated when appropriate according to standard farm protocols.

Two tests for fearfulness were used: novel object and response to human (in this case approach to a stationary, familiar person). These tests were conducted on consecutive days at approximately 41 d of age and repeated at approximately 62 d of age. The response to human test was also conducted at 25 d of age. Tests were conducted between the two daily feedings, but never within 30 min of either feeding time. Novel object tests took place in a test pen that the calves had visited twice daily (for cognitive training; see Meagher et al., 2015) for several weeks. After 2 min of habituation to the pen, the novel object (in this case, a brightly coloured ball) was lowered into the pen using a length of twine. The test lasted 10 min, and latency to make contact with the ball was recorded. The response to human tests were conducted during weekly weighing of the animals, following a similar procedure to Duve et al. (2012) in which calves were allowed to approach a human and then their response to weighing was assessed. In brief, the calf was released from its pen into the alley, and given up to 90 s to make contact with the stationary person. The stationary person (one person per experiment) was familiar to the calves and stood 2.4 m away. The first author (RKM, who was also familiar to the calf) stood inside the pen and recorded the latencies to touch the person. Wooden dividers blocked the view of calves on the other side of the aisle, leaving an alley approximately 1.2 m wide for the individual and pair treatments; however, calves could see into neighbouring pens on the same side of the alley as they approached the person. For

Please cite this article in press as: Meagher, R.K., et al., Inconsistency in dairy calves' responses to tests of fearfulness. Appl. Anim. Behav. Sci. (2016), http://dx.doi.org/10.1016/j.applanim.2016.10.007