# *Evaluating the consistency of dairy goat kids' responses to two methods of assessing fearfulness*

Applied Animal Behaviour Science 272 (2024) 106209

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

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

# Evaluating the consistency of dairy goat kids' responses to two methods of assessing fearfulness

H.M. Vickery [a,*,1,4], F.P. Johansen [b,c,2], R.K. Meagher [d,3]

[a] Department of Animal Sciences, University of Reading, Reading, United Kingdom
[b] School of Biological Sciences, Queens University Belfast, Belfast, Northern Ireland, United Kingdom
[c] Animal Welfare Unit – Livestock Production Sciences Branch, Sustainable Agri-Food Sciences Division, Agri-Food and Biosciences Institute, Hillsborough, United Kingdom
[d] Department of Animal Science and Aquaculture, Dalhousie University, Truro, Nova Scotia, Canada

ARTICLE INFO

*Keywords:*
Behaviour
Welfare
Personality
Goat
Individuality

ABSTRACT

Understanding individual behavioural differences could enhance welfare research. Many methods for assessing fear to make inferences about 'personality' have been proposed but not validated for goat kids. The study's primary aims were to 1) investigate individual-level test-retest reliability of Novel Object (NO) and Familiar Person (FP) tests; 2) establish the effects of testing environment (Modified Home (MH) or Unfamiliar Testing (UT) arenas) and weaning method (Gradually (GW; n=18) or Abruptly (AW; n=17)), and 3) test for associations between repeatable behaviours and production metrics (e.g. growth). A secondary aim was to assess interobserver reliability. One AW pen and one GW pen were used for each testing environment (MH/UT). Four NO tests were conducted, two pre-weaning (25d and 29d), and two post-weaning (63d and 67d); FP tests were conducted the day after each. Tests lasted 180 s (+90 s habituation for NO tests) and were recorded. An observer blinded to weaning treatment scored all videos, and another scored 1 kid's test for each timepoint/pen. One kid from MH and seven from UT environments were removed from testing and subsequent analysis due to distress. Intraclass coefficient calculations indicated good interobserver reliability ($W_1$=0.670, p<0.001). For FP tests 'bipedal stance' ($W_3$=0.379 p<0.001) and 'stand still-look' ($W_3$=0.378 p=0.010) and for NO tests bipedal stance ($W_3$0.234 p=0.006) and 'latency until contact' $W_3$=0.202 p=<0.001 showed test-retest repeatability. MH and UT environments were compared using Kendall's W (calculated for each behaviour separately), the coefficients were compared with Mann-Whitney U tests, which found no impact of testing environment (p=0.579). GLMMs assessed the effect of external factors pen, test type, and weaning treatment; concordance coefficients between individual kids' tests were the dependent variable with test type, kid age and pen as independent variables. The model of best fit was selected on the basis of lowest Akaike Information Criterion (AIC); test type (p=0.184) did not predict individual consistency, but pen did (p=0.002). Further GLMMs found no effect (p>0.05) of 'personality' indicators on weight gain, milk intake/day, nor relationship with weaning treatment, but a significant effect of bipedal stance on milk feeds/day (p=0.04) was identified. Results indicate that 'bipedal stance' in both tests, 'stand still-look' in FP tests, and 'latency to interact' in NO tests have some repeatability, therefore could be 'personality' indicators. Statistical analysis found that testing environment did not affect fear among kids tested, but in practice the unfamiliar environment resulted in more distress-related removals therefore a modified home pen is recommended.

* Corresponding author.
*E-mail address:* holly.vickery@pgr.reading.ac.uk (H.M. Vickery).
[1] https://orcid.org/0000-0001-7533-1136
[2] https://orcid.org/0009-0000-0696-1839
[3] http://orcid.org/0000-0002-6576-8155
[4] Current address: Department of Animal Health, Behaviour and Welfare, Harper Adams University, Edgmond, Newport, United Kingdom.

https://doi.org/10.1016/j.applanim.2024.106209
Received 21 June 2023; Received in revised form 31 January 2024; Accepted 24 February 2024
Available online 28 February 2024

## 1. Introduction

Farm animal research often under-considers the individual in favour of group and pen level approaches (Fraser, 2009), yet animal welfare is defined as 'a measurable quality of living at a specific time for an *individual* animal' (Broom, 2011). Approaches that consider individuality are being increasingly promoted across farm animal welfare literature (Richter and Hintze, 2019; Winckler, 2019). Individuality in behavioural responses of goat kids was reported as early as 1989 (Lyons, 1989), but only recently has the idea that this individuality can be related to consistent 'personality' traits begun to be explored in goats. Further understanding of farm animal personality could help select individuals that can be housed under captive conditions with a lower likelihood of developing welfare problems (Richter and Hintze, 2019).

Fear and its behavioural indicators have been more commonly assessed in the welfare literature without repeated testing to attempt to assess it as a personality trait, and fear is deemed to be of welfare concern due to its experiential negative affective state (Boissy, 1995). This is of particular importance for animals farmed under commercial conditions which require them to adapt to, and cope with, conditions that they did not evolve within (Broom, 2011; Finkemeier et al., 2018), and is of commercial interest due to the heritability of traits such as reactivity and fearfulness (reviewed by Boissy et al., 2005), with links to productivity (Hemsworth et al., 2000). Individuals differ in the level of fear they display in response to a stressor, with more fearful individuals showing stronger behavioural responses (Boissy and Bouissou, 1995), and often decreased productivity (as reviewed in relation to human-animal relationships: Acharya et al., 2022). We are particularly concerned by fear if it is long-lasting or particularly intense, as whilst fleeting, short-lived fear expressed at a level appropriate to a specific stimulus is adaptive, longer lasting or intense fearfulness is maladaptive, particularly in farm environments, and contributes towards stress and an overall impaired welfare state. Therefore, the consistency, and intensity, of fear when measured over time is of interest, and 'absence of general fear' is a heading under the animal welfare principle 'appropriate behaviour' as laid out by the well validated assessment protocols 'Welfare quality' (Botreau et al., 2009).

Presently, fearfulness is typically assessed through response to novelty by utilising novel object tests (e.g., in juvenile goats, Finkemeier et al., 2022; Toinon et al., 2021), and response to humans (e.g., in calves, Bokkers et al., 2009; Meagher et al., 2016) however these tests have been criticised for their limited validation, unclear reliability, and inconsistent methodology (Forkman et al., 2007) that make it difficult to generalise results. Indeed, whilst studies (Chojnacki et al., 2014; Toinon et al., 2021; Winblad von Walter et al., 2021) have assessed the reaction of goat kids to similar tests neither looked at the interobserver reliability or test-retest repeatabilities as these did not relate to their specific aims. There are high variations between goat kids in terms of their early growth rates (Deeming et al., 2016), and evidence from calves has found that more active and exploratory individuals consume solid feed earlier than other calves and have better growth rates (Neave et al., 2018). Whilst there are likely numerous factors that contribute to growth rate differences, individuality or personality has not yet been considered in goat kids. Individuals display a varying ability to cope with stressors and subsequently their likelihood of developing welfare issues differs (Koolhaas et al., 2010), therefore research that aims to understand reliable ways to assess goat kid behavioural responses, and if these meet the criteria to be classified as personality traits, could help us understand other aspects of kids' ability to thrive within commercial conditions.

The term 'personality' should only be used when individuals show differences in their behavioural responses that are consistent both temporally and contextually (Kaiser and Müller, 2021; Stamps and Groothuis, 2010). Whilst in goats there has been some work on personality, this has generally focused on adults with very specific research objectives: it has been demonstrated that adult goats do show stable personality traits that impact other variables (Finkemeier et al., 2022 personality trait boldness affected reversal learning performance: Nawroth et al., 2017 personality trait sociability affected visual discrimination learning). Only one study with kids has reported results from repeated tests that could meet the criteria for personality and found that 61% of 108 goat kids displayed consistent boldness and activity personality traits which were correlated with heart rate variability and weight gain (Finkemeier et al., 2019).

Understanding what traits contribute towards individual variability, and are affected by individuality, may help us to understand how to mitigate the stress of commercial management practices. This seems particularly relevant during the milk feeding stage and weaning transition when goat kids are expected to cope with a combination of stressors including separation from their dams, introduction to a new social and physical environment, and the transition from milk to solid feed without experienced social models. Therefore, the primary aims of the current study were to: 1) investigate individual level test-retest reliability of Novel Object (NO) and Familiar Person (FP) tests; 2) establish the effects of two external factors: testing environment (Modified Home (MH) or Unfamiliar Testing (UT) arenas) to attempt to identify one to recommend for use in further research, and as the study animals were in two different weaning treatments for a concurrent study, abrupt and gradual weaning methods, 3) test for associations between repeatable behaviours that could be personality indicators and production metrics. A secondary aim as part of validating the observational methodology used was to assess interobserver reliability.

## 2. Materials and methods

Ethical approval was granted by the University of Reading, School of Agriculture, Policy and Development (ref. 001561) and Dalhousie University (ref. 2021–010). The goat kids were cared for in accordance with the DEFRA code of recommendations for goats (2013) and the research occurred between June and September 2021.

### 2.1. Animals and housing

The male dairy cross-bred (Saneen, Toggenburg, Alpine) goat kids were used in a concurrent study to describe their feeding behaviour (Vickery et al., 2023a) and compare gradual and abrupt weaning methods (Vickery et al., 2023b); further detail regarding this experimental design is available in Vickery et al. (2023b). Briefly, 35 goat kids were divided into four 3.66 x 3.66 m pens under an Abrupt Weaning treatment (two pens of 8 and 9 kids; AW) that had *ad libitum* access to milk until complete removal at 17:00 hrs on d 56, and Gradual Weaning treatment (two pens of 9 kids; GW) that had *ad libitum* milk until 35 d when milk access was removed for 3.5hrs/d, at 45 d milk was removed for 7hrs/d then full milk removal occurred at 17:00 hrs on d 56.

All kids were cared for according to standard commercial management; bedded on straw topped up daily, castrated via elastration at <7 days of age, vaccinated with clostridial vaccinations at d 21 and 49, and had *ad libitum* access to straw, hay, creep feed and water from d 14. Individual level milk feeding data was recorded from d 22–56 by a computerised milk feeder (Foerster-Technik VARIO smart) supplying one teat per pen, from d 56–60 (the immediate post weaning period) the kids visits to the milk feeding station continued to be monitored. Average Daily Gain (ADG) was calculated based on weekly weighing sessions.

### 2.2. For all testing

One AW pen and one GW pen was used for data collection for each type of testing arena. An external webcam was used to record the tests onto a laptop and the video files were then backed up onto external hard drives. Testing occurred between 15:30 and 19:30hrs. Any kids that exhibited extreme stress behaviours, combining distress vocalizations with more than two escape attempts (jumping into the walls of the test

pen) had the test ended and the kid was immediately returned to the home pen.

*2.2.1. Novel object (NO) test*

These tests were used to assess kid's reaction to novelty as this can be considered a measure of boldness (Frost et al., 2007) and the exploration-avoidance axis (Réale et al., 2007) The novel objects used were four similarly sized but differently coloured and shaped plastic dog toys such that novelty was maintained for all tests, but effects of test order were not confounded with any differences between objects. The toys were disinfected after each use. The kids were tested at age 25 d (yellow frisbee), 29 d (green tube), 63 d (orange wishbone), and 67 d (pink snake). After 90 s of habituation time the novel object was placed inside the pen via a hole cut in the 1.22 m end of the testing corridor, and the test then took place for 180 s.

*2.2.2. Familiar person (FP) test*

As personality traits should be contextually stable (Réale et al., 2007), to provide a contrasting context to novelty kids were tested for their response to a familiar person the day following each of the novel object tests. One kid was released into the testing corridor where a stationary, familiar person (HMV – their daily caregiver) was stood wearing standardised clothing (green overalls) with a neutral expression looking straight ahead at the far end of the corridor (2.44 m away). After 180 s the test ended, and the kids were returned to the group pen.

*2.2.3. Experimental design – environment 1: modified home (MH) pen test*

Pens one (GW nine kids) and two (AW eight kids) were used, two 2.44 m sheeted hurdles were carried into the kids standard 3.66 x 3.66 m home pen one hour before testing began. These sheeted hurdles created an L shaped test arena, for the hour before testing one side was left open so that the kids could familiarise themselves with the testing area. Once the habituation time ended one side of the hurdles was blocked to create a rectangular 2.44×1.22 m corridor. One kid at a time was then lifted over the hurdle into the test pen (a modified familiar environment), leaving all remaining home-pen kids (eight or seven) in the 2.44 x 2.44m pen whilst testing occurred. Kids could hear their pen mates and were also able to perform a bipedal stance against the hurdles to see them (Fig. 1).

*2.2.4. Experimental design – environment 2: unfamiliar testing (UT) arena*

Pens three (GW) and four (AW) were used (nine kids per pen). A 2.44×1.22 m corridor was created within the same barn as the goat kids but outside of their home pens. The corridor was created using 1.22 m tall white hygienic plastic sheets known as 'parlour board' that the kids were unable to see over, but they could hear the goat kids in the same barn (Fig. 1).

*2.3. Behavioural analysis*

The ethogram (Table 1) was developed based on pre-existing literature (Neave et al., 2018; Toinon et al., 2021; Winblad von Walter et al., 2021; Zhang et al., 2021) and initial observations of a subset of the video recordings to ensure all relevant behaviours were captured. All videos were scored using the ethogram (Table 1) in the BORIS software (Friard and Gamba, 2016) by author FPJ who was blinded to the weaning treatments, but due to the visible nature of the different testing methodology (MH/UT) could not be blinded to this factor. Analysis began when the NO was placed inside of the pen, or in the case of the FP test when the kid entered the arena, and the door was closed behind them. A second observer (author HMV) watched and scored 1 kid's test for each timepoint and each pen (a total of 32 videos) in order to test interobserver reliability, however, this observer could not be blinded to weaning treatment due to their involvement in the daily running of the study.

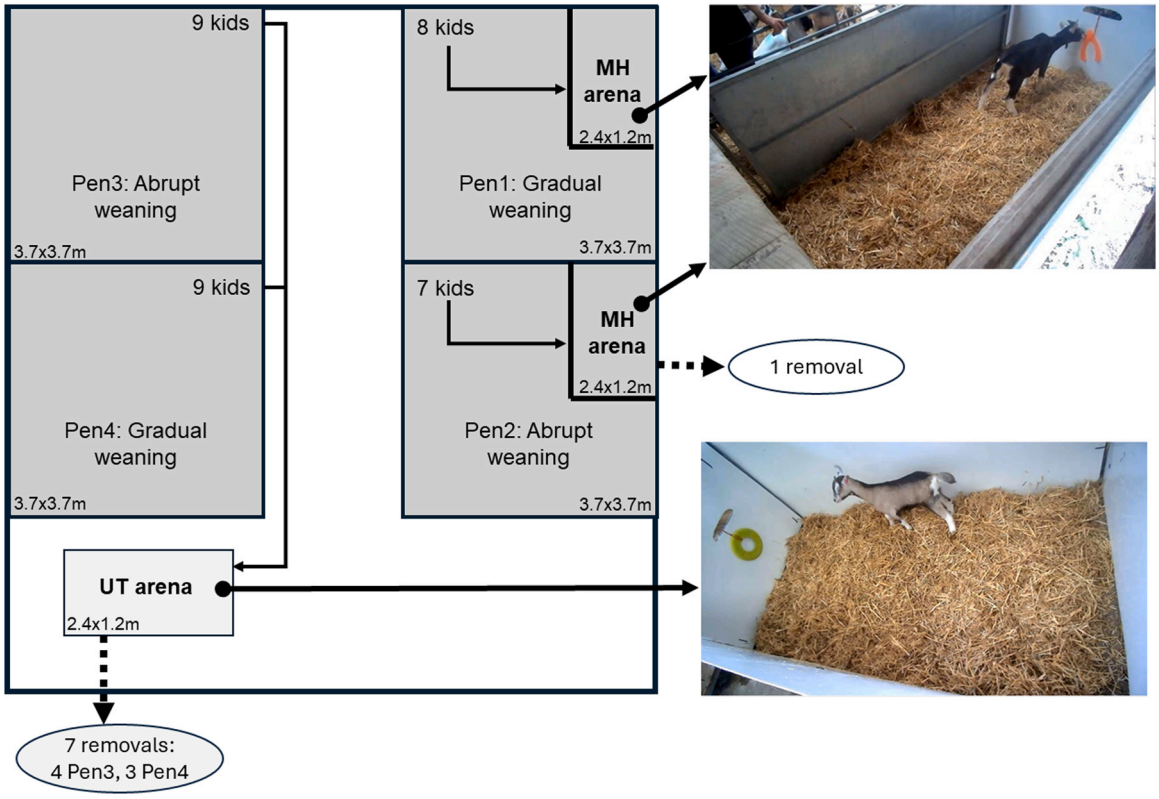


**Fig. 1.** Diagram of the layout of the study barn and testing arenas with arrows indicating kid testing location and removals. UT = Unfamiliar testing; MH = Modified home pen.

**Table 1**
An ethogram of behaviours recorded during novel object and familiar person tests of artificially reared goat kids.

| Behaviour | Type | Description |
|---|---|---|
| Vocalisation | Point | The kid generates and emits noise from its larynx |
| Retreat | Point | Retreat from novel object/familiar person. Backwards steps that begin when looking at/ oriented towards the NO/FP, counted as one point regardless of number of steps until there has been a pause of greater than >5 seconds. |
| Jump | Point | Kid jumps against testing arena wall with all four feet. |
| Sudden movement | Point | A sudden, startle type movement, can be the whole or part of the body including the head or legs. |
| Interact with novel object or familiar person | State | Within 5 cm of the object/human with face orientated towards it – may be sniffing or touching or chewing. |
| Locomotion | State | Movement (such as walking or running but excluding jumping or bipedal stance). Can be in any direction (forwards/backwards) but cannot be stationary movement. Minimum of 2 steps taken. |
| Latency to interact | State | Time in seconds from the novel object being placed inside the testing corridor until the kids' nose is within 5 cm of the novel object/ human |
| Stand still look | State | Standing on with all four feet still looking at the novel object/ familiar person for >5 s, there may be head movements. Timer starts after the 5 seconds have passed. |
| Bipedal stance | State | Kid stands on its back legs with its front legs on the testing corridor wall for >5 s |
| Investigate arena | State | Placing the nose in contact with or less than one muzzle length from the walls or floor of the pen. |
| Rear against pen | Point | Kid jumps against testing arena wall with only front feet, back feet stay in contact with the floor. Differentiated from bipedal stance by momentary (<5 s) nature. |
| Out of view | State | Kid cannot be seen from the camera angle |

### 2.4. Statistical analysis

Two deaths occurred during the study; one from pen 1 (GW, MH environment) from suspected urolithiasis, and one from pen 2 (AW, MH environment) from pneumonia – all their data was excluded from analysis. One kid (AW) from the MH environment, and seven kids (three GW, four AW) from the UT environment were not analysed due to early removal from the tests due to distress. This left 14 kids tested in the MH environment (6 AW; 8 GW) and 11 kids tested in the UT environment (5 AW; 6 GW). When NO testing, each kid had 90 s to acclimatise to the testing pen. During this time, they had free access to the pen, and therefore when the NO was presented kids were located in different pen areas. To control for the effects of these different start locations on the kids' subsequent latency to interact with the object, the pen was visually divided into three equal sections (A – closest to the novel object, B – middle of pen, C – furthest from novel object). The videos were analysed to find kids in the different sections that immediately approached the novel object. These kids were timed, and the values were averaged per section. The variable 'Latency to interact with novel object' was then adjusted to control for where the kid started their test. Calculations were performed using the following formulas: Section A = latency + average value for section C, Section B = latency + (average value for section C – average value for section B), Section C = no change. Unless otherwise specified statistical analysis was conducted in IBM SPSS Statistics (Version 27).

#### 2.4.1. Test-retest reliability – assessment of possible personality indicators

Kendall's test of concordance was carried out for each behavioural variable. Testing occasions were compared across time (tests 1–4) for the NO test and the FH test. Behaviours found to be concordant ($W<0.2$) were averaged across occasions to create one value per kid, these values were then retained for further analysis. The cut-off selected was 0.2, which is relatively low but was chosen due to the novelty of the study and the low sample size. If the same behaviour was found to be concordant between tests (NO/FH), and positively correlated, the behaviour was averaged between tests. Variables that were found to not be concordant were discarded. Spearman's rank correlation was carried out to assess the relationship between 'Bipedal stance' in familiar human and in novel object tests.

#### 2.4.2. Influence of external factors

Consistency is considered a key component of personality, therefore behavioural tests which elicit individually consistent behaviours can be considered 'good' tests of personality. To fulfil aim 3 and investigate behavioural differences between the two testing arena types (Modified home pen vs unfamiliar), the two arena types were compared using Kendall's W. These were calculated for every behaviour of each treatment group separately (pens 1 and 2; pens 3 and 4) and the resulting coefficients for each behaviour were compared between groups with Mann-Whitney U tests.

To assess the effects of external factors on the consistency of kid responses to personality testing, two GLMMs were used. Both models included kid ID as a random effect and Kendall's W for each kid across testing types as the dependent variable. Model 1 included test type, pen, and the interaction between test type and pen as independent variables, and model 2 included test type, weaning method, and the interaction between test type and weaning method as independent variables. Separate models were run as weaning method and pen showed high multicollinearity. The best fitting model for each dependent variable was selected on the basis of lowest Akaike Information Criterion (AIC). Model residuals were checked for normality and homoscedasticity using visual assessment of Q-Q and residuals vs predicted value plots.

#### 2.4.3. Association of repeatable behaviours and production metrics

Variables were tested for normality and transformed where necessary. In line with aims 4 and 5, a series of GLMMs were conducted to assess the correlation between repeatable behaviours that could indicate personality (as the independent variables) and production metrics (as the dependant variable.). Model structure was as follows; random effect of kid nested within pen; dependent variables: average daily gain, mean daily milk intake, mean number of milk feeds per day (square root transformed), day kid learnt to use the milk feeder independently, weaning treatment, and mean number of visits to the milk feeder post-weaning; independent variables: bipedal stance (average across FH/NO, square root transformed), Stand still-look (FH, log10 transformed), Latency to touch novel object (LN transformed). Models were checked and tested as in section 2.5.2.

#### 2.4.4. Inter-observer reliability

Intraclass correlations (ICC) estimates along with 95% confidence intervals were calculated in RStudio version 4.3.0 (RStudio, 2021) using the "psych" package (Revelle, 2024). Tests were based on a single-rating, absolute agreement, two-way random-effects model (ICC (2,1)). ICC tests were carried out on raw duration data (s) which was transformed into proportional data (%) for further analysis.

## 3. Results

### 3.1. Test-retest reliability – assessment of possible personality indicators

Results indicate that 'Bipedal stance' in both novel object and familiar human tests, 'stand still look' in the familiar human tests, and 'Latency until contact' in the novel object tests are repeatable personality measures for male goat kids (Table 2). There was a positive

**Table 2**
Results of Kendall's W tests for each behaviour in familiar person and novel object tests.

| Behaviour | Familiar person test | | | Novel object test | | |
|---|---|---|---|---|---|---|
| | N | Kendall's W | Significance | N | Kendalls' W | Significance |
| Bipedal stance | 18 | 0.379* | <0.001 | 18 | 0.234* | 0.006 |
| Interact with NO/FH | 28 | 0.124 | 0.016 | 28 | 0.012 | 0.797 |
| Investigate arena | 29 | 0.037 | 0.356 | 29 | 0.027 | 0.502 |
| Jump | 24 | 0.124 | 0.031 | 24 | 0.017 | 0.739 |
| Latency until contact | 29 | 0.030 | 0.462 | 29 | 0.202* | <0.001 |
| Locomotion | 29 | 0.022 | 0.586 | 29 | 0.132 | 0.010 |
| Out of view | 29 | 0.003 | 0.973 | 22 | 0.049 | 0.354 |
| Stand still look | 10 | 0.378* | 0.010 | 10 | 0.115 | 0.328 |
| Vocalisation | 29 | 0.132 | 0.010 | 29 | 0.105 | 0.027 |

Kendall's W values >0.200 are marked with a *.

correlation between 'Bipedal stance' in FH and NO tests (r(26) =0.81, p>0.001), so these were averaged into 'Bipedal stance all' for further analysis. 'Stand still look' – familiar person test and 'Latency until contact' – novel object test, were also retained for further analysis.

### 3.2. Influence of external factors

Results of Kendall's W tests showing the differences between testing environments can be seen in Table 3; the subsequent Mann-Whitney U test found no significant difference between the concordance coefficients of the two testing environments (p=0.579).

**Model 1**. There were no effects of test type or test type*pen on kid consistency. There was, however, an effect of pen (F(51) =4.399, p=0.008). Post-hoc tests found that there were significant differences in kid consistency between pens 1 + 3 (p=0.005) and pens 2 + 3 (p=0.012). Kids in pens 1 and 2 (MH environment) were generally less consistent across testing occasions (Fig. 2).

**Model 2**. There were no significant effects of treatment, test type, or treatment*test type on kid consistency.

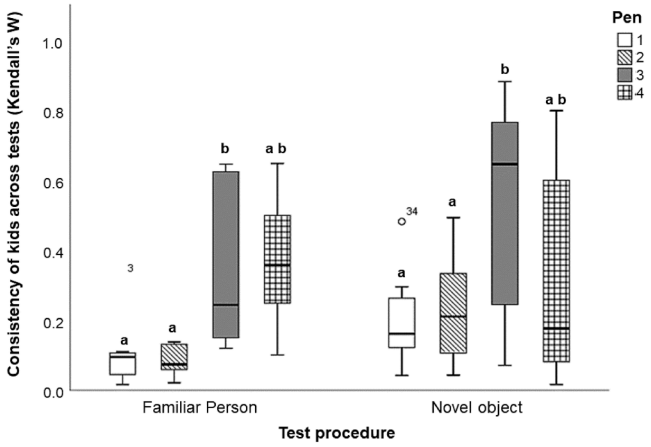


**Fig. 2.** A boxplot showing differences in Kendall's W values between pens of goat kids (Pens 1 and 2 = modified home pen environment; Pens 3 and 4 = unfamiliar testing environment) and test procedure.

**Table 3**
Results of Kendall's W tests for each behaviour in novel object and familiar person tests for each testing environment.

| Test | Behaviour | Modified home pen | | | Unfamiliar arena | | |
|---|---|---|---|---|---|---|---|
| | | N | Kendall's W | Significance | N | Kendall's W | Significance |
| **Novel object** | Bipedal stance | 15 | 0.313* | 0.003 | 3 | N/A | |
| | Interact with NO/FH | 14 | 0.051 | 0.543 | 14 | 0.099 | 0.247 |
| | Investigate arena | 14 | 0.098 | 0.25 | 14 | 0.055 | 0.510 |
| | Jump | 13 | 0.026 | 0.794 | 11 | 0.040 | 0.727 |
| | Latency until contact | 15 | 0.291* | 0.004 | 14 | 0.166 | 0.073 |
| | Locomotion | 15 | 0.035 | 0.665 | 14 | 0.334* | 0.003 |
| | Out of view | 15 | 0.188 | 0.037 | 7 | 0.184 | 0.277 |
| | Stand still look | 5 | 0.253* | 0.284 | 5 | 0.133 | 0.572 |
| | Vocalisation | 15 | 0.125 | 0.133 | 14 | 0.158 | 0.084 |
| **Familiar person** | Bipedal stance | 15 | 0.513* | <0.001 | 3 | N/A | N/A |
| | Interact with NO/FH | 14 | 0.096 | 0.258 | 14 | 0.376* | 0.001 |
| | Investigate arena | 15 | 0.063 | 0.418 | 15 | 0.130 | 0.118 |
| | Jump | 13 | 0.183 | 0.068 | 11 | 0.145 | 0.189 |
| | Latency until contact | 15 | 0.154 | 0.074 | 14 | 0.059 | 0.481 |
| | Locomotion | 15 | 0.068 | 0.379 | 15 | 0.07 | 0.372 |
| | Out of view | 15 | 0.050 | 0.522 | 7 | N/A | N/A |
| | Stand still look | 5 | 0.244* | 0.300 | 5 | 0.600* | 0.029 |
| | Vocalisation | 15 | 0.298* | 0.004 | 15 | 0.169 | 0.055 |

Kendall's W values >0.200 are marked with a *. N/A values are presented where the sample size was too low to complete the test.

### 3.3. Association of repeatable behaviours and production metrics

There was a significant association between latency to interact with the novel object, and average number of daily milk feeder visits post-weaning: an increase in average number of feeder visits was associated with a 0.688 second decrease in latency to interact with the novel object (t=-2.077, p=0.049, 95% CI= {-1.372, –0.004}) (Fig. 3). There was also a significant association between average number of milk feeds per day and average proportion of tests spent in a bipedal stance: an increase in milk feeds per day was associated with kids spending 1.72% longer in a bipedal stance (t=2.156, p=0.04, 95% CI= {0.08, 3.36}) (Fig. 4). No relationship was found between ADG, average daily milk intake, weaning treatment, and day kid learned to use the feeder independently and any of the repeatable behaviours.

### 3.4. Inter-observer reliability

ICC tests indicated good reliability (ICC=0.87, (n=2, df=433) p<0.001, 95% CI {0.84–0.89})

## 4. Discussion

The methodologies described here elicited good interobserver reliability, which is an important aspect of ensuring repeatability. Out of the nine behaviours recorded, three ('bipedal stance' in both tests, 'stand still-look' in FP tests, and 'latency to interact' in NO tests) showed test-retest repeatability, therefore could be considered 'personality' indicators. Testing in an unfamiliar environment or modified home pen did not impact behavioural results when kids completed their tests, but more kids were removed from testing in the former due to distress. When investigating individual kid consistency, test type did not predict this, but the pen the kid was housed in did; however, due to a low number of both individual kid and pen-level replicates, distinguishing this relationship is difficult. No relationship was found between the repeatable behaviours (considered 'personality' indicators) and weight gain, daily milk intake, or weaning treatment, but a trend towards 'latency to interact with NO' predicting average milk station visits post-weaning, and a significant relationship between bipedal stance and daily milk feeds were identified.

Home pen tests are desirable for practical and welfare reasons because they avoid introducing handling effects and complete social isolation (see Forkman et al., 2007; Tecott and Nestler, 2004 for further discussion). In our case, home pen modification was less feasible (carrying equipment into the home pen was practically challenging) than the unfamiliar environment, hence both were utilised to test for differences in the measured behavioural responses – none were found,

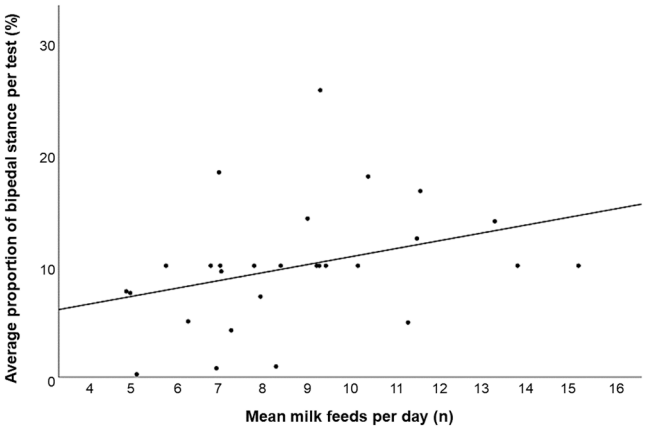


**Fig. 3.** Relationship between the behaviour 'bipedal stance' performed by goat kids in novel object and familiar human tests and the mean number of milk feeds they fed in per day (R2 = 0.118, y=0.04+7.09E-3*x).

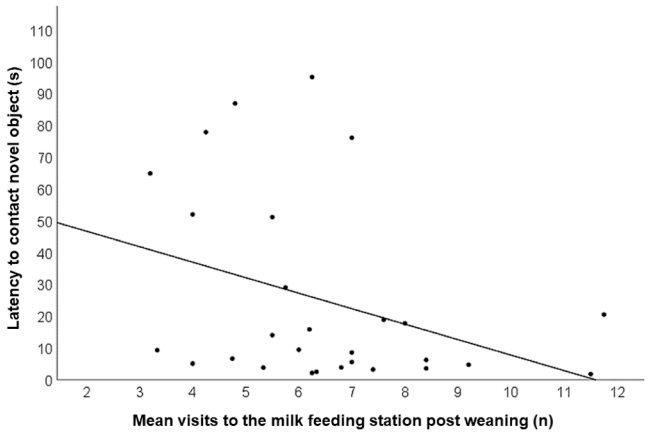


**Fig. 4.** Relationship between the behaviour 'latency to contact' the novel object and mean number of visits to the milk feeding station post weaning of goat kids reared artificially (R2 = 0.122, y=56.54–4.88*x).

implying that either could be used (depending on practicality). Modified home pens reduce social isolation (our kids were able to perform bipedal stance and see their pen-mates), and isolated fear tests have received criticism due to lacking species-appropriate consideration of gregariousness (Forkman et al., 2007). Whilst this seemingly did not influence our results, it likely affected those with the strongest behavioural responses to isolation, hence more kids being removed from the unfamiliar environment which had complete visual social isolation. Eliminating kids with strong behavioural responses from testing and subsequent analysis removes valuable information about behavioural reactivity but was necessary to protect welfare. Therefore, we suggest that the validation of testing methodologies that negate the need for social isolation, (Toinon et al., 2021 reported results of group-level testing of goat kids, however they did not assess repeatability, so the validity remains unclear) whilst still accurately assessing individual responses is an important area for further research – particularly as welfare relevance would be improved (kids are most likely to experience stressful management changes whilst in a group).

Although our study found no effect of testing environment on kid response, there was a pen-level effect which suggests a possible social factor related to group housing composition that impacts kids' responses. Whilst caution must be taken with regards to these results due to low number of pen replicates in this study, recently, increasing attention has been paid to the effects of gut microbiota on farm animal health and behaviour (comprehensive reviews: Chen et al., 2021; Kraimi et al., 2018), and to its role as a driver of individual behavioural variation (reviewed by Davidson et al., 2018). Our kids were housed in four pens with solid sides, with pens one and two, and pens three and four next to each other (see Fig. 1) such that they may have been able to interact through joins in the pen; perhaps shared pen-level differences in microbiota (a 'social microbiome' being the microbial metacommunity of an animal social group: Sarkar et al., 2020) could explain some of the pen effect identified. We suggest that considering the gut microbiome research may be valuable to understanding individuality, and that aiming for a greater number of pen replicates would be desirable.

Meagher et al. (2016) hypothesised that low reliability in one of their calf experiments was due to a long test-retest interval (20 days) and that milk weaning occurred during the interval; our study chose to repeat tests with a four-day interval both pre and post weaning to avoid this possible issue, however differences in consistency between the tests pre- and post-weaning were not observed. A further issue with response-based tests that measure voluntary approaches is that individuals who are simply 'indifferent' to the object or human may show similar behavioural responses, such as a long latency to approach, as kids that are truly fearful (discussed for mink in Meagher et al., 2011).

Understanding and measuring additional behavioural responses which can differentiate between these kids is important, and vocalisations could provide further information on this. Whilst we did record vocalisations, we didn't attempt to assess differences between them as we felt this would not be reliable or repeatable without full acoustic analysis. We would highly recommend this be combined into future work as we anecdotally observed an apparent difference between shorter contact-type calls, and louder, longer, seemingly more distressed calls, a difference which has been found in previous acoustic analyses (Siebert et al., 2011).

Personality traits are inter-individual behavioural differences that are *both* temporally and contextually consistent (Koolhaas et al., 2010; Réale et al., 2007; Wolf and Weissing, 2012), and there is increasing evidence that they are related to growth and productivity of ruminants (Haskell et al., 2014; Neave et al., 2018). Whilst our study focused largely on indicators of fearfulness, which is not explicitly considered a dimension of personality by Finkemeier et al., (2018) (they discuss exploration, sociability, aggressiveness, boldness and activity), the behaviours observed here link to exploration, sociability and boldness. Furthermore, fearfulness as explored by the tests we have used has been a key interest of farm animal 'temperament' research (see Forkman et al., 2007 for a review) which Finkemeier et al., (2018) states is an aspect of, and highly interrelated with personality. Nonetheless, our study identified just three repeatable behaviours (meaning a principal component analysis typically used for these kinds of analyses could not be employed) and only two associations were found between repeatable behaviours (possible personality traits) and production metrics. An association between latency to interact with the novel object, and the average number of daily milk feeder visits kids made postweaning was identified, with kids visiting the milk feed station more having a lower latency to interact with the novel object. This is likely to reflect that these kids showed greater levels of activity and general 'boldness' (a personality trait that has been identified by principal component analysis in juvenile goats; Finkemeier et al., 2019), which could be linked to a more pro-active style of coping with the management change of weaning. It may also reflect a generally higher level of neophilia with a desire for novelty reflected by more frequently visiting the slightly different environment of the milk feeding station, indeed Nawroth et al., (2017) classified latency to interact within an 'explorative' personality trait category.

A further association between average number of daily milk feeds and average proportion of tests spent in a bipedal stance was identified, higher daily milk feeds were associated with kids spending longer in a bipedal stance during testing. Bipedal stance could reflect both exploratory and social personality traits (Nawroth et al., 2017) if we consider that they were likely performing this behaviour to widen their field of view, perhaps in response to isolation (to visually search for their conspecifics), or to more generally look around and explore the environment, but there is little goat-specific research to inform our interpretation of its association with milk feeding. Dairy cow research has identified a link between alertness and feed consumption with cows that were more alert/ neophilic consuming more, and more fearful animals consuming less (Schwanke et al., 2022), and calves that show higher exploration in novel object tests are faster to consume concentrate feed, have higher growth and intakes (Neave et al., 2018). Whilst bipedal stance could be indicative of alertness and exploration its links to feeding are unclear as there was no association found with overall milk intake or growth, but only how many feeds they consumed that milk in. However, our study was limited by an inability to record individual solid feed intakes, which prevented further exploring links with solid feeding behaviour. Developing our understanding of behavioural individuality could have practical implications; in cattle, personality traits have been shown to have moderate to high levels or heritability (Haskell et al., 2014; Stirling et al., 2002) which could influence selective breeding programmes. Furthermore, individuals less suited to living under commercial conditions are likely to experience poorer welfare, which is of concern for reasons of ethics and associations with reduced productivity (Voisinet et al., 1997; Burrow and Dillon, 1997).

The study size needed to assess personality correlation estimates accurately has been proposed as 250 individuals (Schönbrodt and Perugini, 2013), which with limited research funding and farm sizes is realistically impossible to achieve in most individual studies; indeed these practical constraints limited both our individual kid and pen-level replicates and influence the level of confidence we can have in our results. As per Meagher et al., (2017)'s recommendation we agree that multi-study replication with meta-analyses should be the gold standard and hope that the information reported in our study regarding testing methodologies is a step towards a standardised protocol. With the recent establishment of the 'Many Goats' project (www.themanygoatsproject.com) following on from the success of multi-lab collaborations including the 'Many babies' project (Frank, 2016), further work on goat personality with larger sample sizes should be forthcoming, but we stress that this should consider all life stages and not just adults.

### 4.1. Conclusions

The behavioural ethogram and testing procedures used elicited good interobserver reliability, and it was found that test type did not predict individual consistency, but pen did, which warrants further investigation. In kids tested repeatedly before ten weeks old only three out of nine behaviours recorded showed repeatability which suggests they could be indicators of personality; 'bipedal stance' in both tests, 'stand still-look' in FP tests, and 'latency to interact' in NO tests. There were few links between the personality indicators identified and production metrics; no correlations were found with weight gain, milk intake/day, or weaning treatment, but a trend towards effect of 'latency to interact with NO' on average milk station visits post-weaning (shorter latency to interact = more visits to the milk station), and an effect of bipedal stance on milk feeds/day (more time in bipedal stance = more feeds) were found. Whilst the results of this small-scale study of artificially reared male goat kids should be interpreted with caution due to limited individual and pen-level replicates, it was found that testing environment did not statistically affect fear in kids that completed their tests, but the unfamiliar environment did result in more distress-related removals therefore a modified home pen environment is recommended wherever practical, and several areas for further investigation have been identified.

## Ethics approval

Ethical approval was granted by the University of Reading, School of Agriculture, Policy and Development, ref. 001561 and Dalhousie University, ref. 2021–010.

## Financial support statement

Financial support was provided by the School of Agriculture, Policy and Development, University of Reading, the West Country Dairy Awards (Charity no. 306598), and the Perry foundation (Charity no. 310885).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Our appreciation goes to Volac International and Forster-Technik for loaning the milk feeder for the second project and to Will Sinclair for his expertise and technical support, as well as to the Vickery family for hosting and supporting this research.

## References


Acharya, R.Y., Hemsworth, P.H., Coleman, G.J., Kinder, J.E., 2022. The animal-human interface in farm animal production: animal fear, stress, reproduction and welfare. Animals 12, 487. https://doi.org/10.3390/ani12040487.

Boissy, A., 1995. Fear and fearfulness in animals. Q. Rev. Biol. 70, 165–191. https://doi.org/10.1086/418981.

Boissy, A., Bouissou, M.-F., 1995. Assessment of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations. Appl. Anim. Behav. Sci. 46, 17–31. https://doi.org/10.1016/0168-1591(95)00633-8.

Boissy, A., Fisher, A.D., Bouix, J., Hinch, G.N., Le Neindre, P., 2005. Genetics of fear in ruminant livestock. Livest. Prod. Sci., Genet. Behav. 93, 23–32. https://doi.org/10.1016/j.livprodsci.2004.11.003.

Bokkers, E.A.M., Leruste, H., Heutinck, L.F.M., Wolthuis-Fillerup, M., Werf, J. van der, Lensink, B.J., Reenen, C. van, 2009. Inter-observer and test-retest reliability of on-farm behavioural observations in veal calves. Anim. Welf. 18, 381–390. https://doi.org/10.1017/S0962728600000786.

Botreau, R., Veissier, I., Perny, P., 2009. Overall assessment of animal welfare: strategy adopted in Welfare Quality®. Anim. Welf. 18, 363–370. https://doi.org/10.1017/S0962728600000762.

Broom, D.M., 2011. A history of animal welfare science. Acta Biotheor. 59, 121–137. https://doi.org/10.1007/s10441-011-9123-3.

Burrow, H.M., Dillon, R.D., 1997. Relationship between temperament and growth in a feedlot and commercial carcass traits in Bos indicus crossbreeds. Aust. J. Exp. Agric. 37, 407–411. https://doi.org/10.1071/EA96148.

Chen, S., Luo, S., Yan, C., 2021. Gut microbiota implications for health and welfare in farm animals: a review. Anim. (Basel) 12, 93. https://doi.org/10.3390/ani12010093.

Chojnacki, R.M., Vas, J., Andersen, I.L., 2014. The effects of prenatal stocking densities on the fear responses and sociality of goat (Capra hircus) kids. PLoS One 9, e94253. https://doi.org/10.1371/journal.pone.0094253.

Davidson, G.L., Cooke, A.C., Johnson, C.N., Quinn, J.L., 2018. The gut microbiome as a driver of individual variation in cognition and functional behaviour. Philos. Trans. R. Soc. B: Biol. Sci. 373, 20170286. https://doi.org/10.1098/rstb.2017.0286.

Deeming, L., Beausoleil, N., KJ, S., Webster, J., Zobel, G., 2016. Variability in growth rates of goat kids on 16 New Zealand dairy goat farms. Proceedings of the New Zealand Society of Animal Production 76, 137–138.

Finkemeier, M.-A., Krause, A., Tuchscherer, A., Puppe, B., Langbein, J., 2022. Personality traits affect learning performance in dwarf goats (Capra hircus). Front. Vet. Sci. 9.

Finkemeier, M.-A., Langbein, J., Puppe, B., 2018. Personality research in mammalian farm animals: concepts, measures, and relationship to welfare. Front. Vet. Sci. 5.

Finkemeier, M.-A., Oesterwind, S., Nürnberg, G., Puppe, B., Langbein, J., 2019. Assessment of personality types in Nigerian dwarf goats (Capra hircus) and cross-context correlations to behavioural and physiological responses. Appl. Anim. Behav. Sci. 217, 28–35. https://doi.org/10.1016/j.applanim.2019.05.004.

Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E., Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiol. Behav., Stress Welf. Farm Anim. 92, 340–374. https://doi.org/10.1016/j.physbeh.2007.03.016.

Frank, M.C., 2016. A collaborative approach to infant research: promoting reproducibility, best practices, and theory-building. Infancy 22 (4), 421–435.

Fraser, D., 2009. Animal behaviour, animal welfare and the scientific study of affect. Appl. Anim. Behav. Sci., Spec. Issue.: Anim. Suff. Welf. 118, 108–117. https://doi.org/10.1016/j.applanim.2009.02.020.

Friard, O., Gamba, M., 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol. Evol. 7, 1325–1330. https://doi.org/10.1111/2041-210X.12584.

Frost, A.J., Winrow-Gifen, A., Ashley, P.J., Sneddon, L.U., 2007. Plasticity in animal personality traits: does prior experience alter the degree of boldness? Proc. R. Soc. B: Biol. Sci. 274, 333–339.

Haskell, M.J., Simm, G., Turner, S.P., 2014. Genetic selection for temperament traits in dairy and beef cattle. Front. Genet. 5.

Hemsworth, P.H., Coleman, G.J., Barnett, J.L., Borg, S., 2000. Relationships between human-animal interactions and productivity of commercial dairy cows. J. Anim. Sci. 78, 2821–2831. https://doi.org/10.2527/2000.78112821x.

Kaiser, M.I., Müller, C., 2021. What is an animal personality? Biol. Philos. 36 (1) https://doi.org/10.1007/s10539-020-09776-w.

Koolhaas, J.M., de Boer, S.F., Coppens, C.M., Buwalda, B., 2010. Neuroendocrinology of coping styles: towards understanding the biology of individual variation. Front. Neuroendocrinol. 31, 307–321. https://doi.org/10.1016/j.yfrne.2010.04.001.

Kraimi, N., Calandreau, L., Biesse, M., Rabot, S., Guitton, E., Velge, P., Leterrier, C., 2018. Absence of gut microbiota reduces emotional reactivity in Japanese Quails (Coturnix japonica). Front. Physiol. 9.

Lyons, D.M., 1989. Individual differences in temperament of dairy goats and the inhibition of milk ejection. Appl. Anim. Behav. Sci. 22, 269–282. https://doi.org/10.1016/0168-1591(89)90022-1.

Meagher, R.K., Duncan, I., Bechard, A., Mason, G.J., 2011. Who's afraid of the big bad glove? Testing for fear and its correlates in mink. Appl. Anim. Behav. Sci. 133, 254–264. https://doi.org/10.1016/j.applanim.2011.05.009.

Meagher, R.K., von Keyserlingk, M.A.G., Atkinson, D., Weary, D.M., 2016. Inconsistency in dairy calves' responses to tests of fearfulness. Appl. Anim. Behav. Sci. 185, 15–22. https://doi.org/10.1016/j.applanim.2016.10.007.

Nawroth, C., Prentice, P.M., McElligott, A.G., 2017. Individual personality differences in goats predict their performance in visual learning and non-associative cognitive tasks. Behav. Process. 134, 43–53. https://doi.org/10.1016/j.beproc.2016.08.001.

Neave, H.W., Costa, J.H.C., Weary, D.M., von Keyserlingk, M.A.G., 2018. Personality is associated with feeding behavior and performance in dairy calves. J. Dairy Sci. 101, 7437–7449. https://doi.org/10.3168/jds.2017-14248.

RStudio Team., 2021. RStudio: Integrated Development. Environment for R. RStudio, PBC, Boston, MA.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010x.

Revelle, W., 2024. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.4.1.

Richter, S.H., Hintze, S., 2019. From the individual to the population – and back again? Emphasising the role of the individual in animal welfare science. Appl. Anim. Behav. Sci. 212, 1–8. https://doi.org/10.1016/j.applanim.2018.12.012.

Sarkar, A., Harty, S., Johnson, K.V.-A., Moeller, A.H., Archie, E.A., Schell, L.D., Carmody, R.N., Clutton-Brock, T.H., Dunbar, R.I.M., Burnet, P.W.J., 2020. Microbial transmission in animal social networks and the social microbiome. Nat. Ecol. Evol. 4, 1020–1035. https://doi.org/10.1038/s41559-020-1220-8.

Schönbrodt, F.D., Perugini, M., 2013. At what sample size do correlations stabilize? J. Res. Personal. 47, 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Schwanke, A.J., Dancy, K.M., Neave, H.W., Penner, G.B., Bergeron, R., DeVries, T.J., 2022. Effects of concentrate allowance and individual dairy cow personality traits on behavior and production of dairy cows milked in a free-traffic automated milking system. J. Dairy Sci. 105, 6290–6306. https://doi.org/10.3168/jds.2021-21657.

Siebert, K., Langbein, J., Schön, P.-C., Tuchscherer, A., Puppe, B., 2011. Degree of social isolation affects behavioural and vocal response patterns in dwarf goats (Capra hircus). Appl. Anim. Behav. Sci. 131, 53–62. https://doi.org/10.1016/j.applanim.2011.01.003.

Stamps, J., Groothuis, T.G.G., 2010. The development of animal personality: relevance, concepts and perspectives. Biol. Rev. 85, 301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x.

Stirling, D.G., Réale, D., Roff, D.,A., 2002. Selection, structure and the heritability of behaviour. J. Evolut. Biol. 15, 277–289. https://doi.org/10.1046/j.1420-9101.2002.00389.x.

Tecott, L.H., Nestler, E.J., 2004. Neurobehavioral assessment in the information age. Nat. Neurosci. 7, 462–466. https://doi.org/10.1038/nn1225.

Toinon, C., Waiblinger, S., Rault, J.-L., 2021. Maternal deprivation affects goat kids' stress coping behaviour. Physiol. Behav. 239, 113494 https://doi.org/10.1016/j.physbeh.2021.113494.

Vickery, H.M., Meagher, R.K., Stergiadis, S., Neal, R.A., 2023a. A preliminary investigation of the feeding behaviour of dairy goat kids reared away from their dams on a computerised ad libitum milk feeding system. Appl. Anim. Behav. Sci. 261, 105898 https://doi.org/10.1016/j.applanim.2023.105898.

Vickery, H.M., Neal, R.A., Stergiadis, S., Meagher, R.K., 2023b. Gradually weaning goat kids may improve weight gains whilst reducing weaning stress and increasing creep feed intakes. Front. Vet. Sci. 10 https://doi.org/10.3389/fvets.2023.1200849.

Voisinet, B.,D., Grandin, T., Tatum, J.,D., O'Connor, S.,F., Struthers, J.,J., 1997. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. J. Anim. Sci. 75, 892–896.

Winblad von Walter, L., Forkman, B., Hogberg, M., Hydbring-Sandberg, E., 2021. The effect of mother goat presence during rearing on kids' response to isolation and to an arena test. Animals 11, 575. https://doi.org/10.3390/ani11020575.

Winckler, C., 2019. Assessing animal welfare at the farm level: do we care sufficiently about the individual? Anim. Welf. 28, 77–82. https://doi.org/10.7120/09627286.28.1.077.

Wolf, M., Weissing, F.J., 2012. Animal personalities: consequences for ecology and evolution. Trends Ecol. Evol. 27, 452–461. https://doi.org/10.1016/j.tree.2012.05.001.

Zhang, C., Juniper, D.T., Meagher, R.K., 2021. Effects of physical enrichment items and social housing on calves' growth, behaviour and response to novelty. Appl. Anim. Behav. Sci. 237, 105295 https://doi.org/10.1016/j.applanim.2021.105295.