Applied Animal Behaviour Science 265 (2023) 105999

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

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

Measuring the human–animal relationship in cows by avoidance distance at pasture

L. Aubé[a,b,1,*], E. Mollaret[a,1], M.M. Mialon[b], L. Mounier[a,b], I. Veissier[b], A. de Boyer des Roches[a,b]

[a] Chaire Bien-Être Animal, VetAgro Sup, 1 avenue Bourgelat, 69280 Marcy-l'Etoile, France
[b] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, 63122 Saint-Genès-Champanelle, France

A R T I C L E I N F O

Keywords:
Dairy cattle
Welfare
Human–animal relationship
Reliability
Pasture

A B S T R A C T

The human–animal relationship is an essential part of farm animal welfare. Avoidance distance at the feeding rack (**ADF**) has been validated as a measure of human–animal relationship (**HAR**) for cows indoors, but no measure has been validated for cows at pasture yet. The aim of this study was to test the validity of 'avoidance distance at pasture' (**ADP**) as a measure of HAR for grazing dairy cows. We assessed the following validity criteria for both ADP and ADF: inter-observer reliability, test–retest reliability, and stability over the grazing season. We also assessed the selectivity of ADP by convergent validity with ADF. Two persons (randomly alternating between the experimenter carrying out the test and the observer watching the test) estimated ADF and ADP twice on 48 dairy cows. During the grazing season, one person measured both ADF and ADP on the 48 cows two days apart for test–retest reliability and 7 times (at 5-week intervals) for stability over the season. For these tests, the experimenter approached the cow from the front, with his hand held at approximately 45°, and the distance between the experimenter's hand and the cow was measured at the moment of the cow's first avoidance movement. The inter-observer reliability, test–retest reliability, and stability over the season of ADF and ADP were analyzed at individual level using both reliability (e.g. Spearman's rank correlation (**RS**), Kendall's coefficient of concordance (**KW**)), and agreement parameters (e.g. smallest detectable change (**SDC**)). A linear mixed model was used to investigate the relationship between ADF and ADP. Inter-observer reliability was high (i.e. satisfactory) for ADF and ADP (RS>0.7). SDC was around 55 cm for ADF whereas it was 127 cm for ADP. Test–retest reliability was high for ADF (RS=0.7) and moderate for ADP (RS=0.5). Stability over the season was moderate for both ADF and ADP (0.5<KW<0.6). We found no significant relationship between ADF and ADP. In conclusion, the ADP measure has acceptable inter-observer reliability but is not stable over short-term or long-term time and is not related to ADF. ADP cannot be entirely validated as a measure of HAR at this stage. Further studies are warranted to address factors that make ADP variable with time and to investigate the selectivity of this measure through a construct validity process.

1. Introduction

The human–animal relationship (**HAR**) is a dynamic process that results from previous human–animal interactions and influences the way that both the human and the animal perceive future interactions (Coleman et al., 1998; Waiblinger et al., 2006, 2003). A poor HAR can have negative effects on the animal's emotional and physiological states (e.g. stress levels, reproductive performances) (Hemsworth, 2003; Napolitano et al., 2020; Estep and Hetts, 1992 cited by Waiblinger et al., 2006). Conversely, a good HAR is beneficial for both animals and humans, for example by reducing stress during handling and therefore accidents (Waiblinger et al., 2002). Measuring HAR is therefore an essential component of animal welfare assessment protocols.

Various measures have been developed to assess HAR in dairy cattle (see Waiblinger et al., 2006, for a review). One of these measures is avoidance distance, which is the distance at which a human can approach an immobile animal before it moves (Battini et al., 2011). Avoidance distance at the feeding rack (**ADF**) is usually used to measure

* Corresponding author at: Chaire Bien-Être Animal, VetAgro Sup, 1 avenue Bourgelat, 69280 Marcy-l.625*pt*′Etoile, France.
E-mail addresses: lydiane.aube@gmail.com (L. Aubé), est.mollaret@gmail.com (E. Mollaret).

[1] Joint first authors (co-first authorship)

https://doi.org/10.1016/j.applanim.2023.105999
Received 13 February 2023; Received in revised form 15 June 2023; Accepted 27 June 2023
Available online 30 June 2023

HAR for animals indoors (e.g. in the Welfare Quality Consortium, 2009; Waiblinger et al., 2003). Avoidance distance at pasture (**ADP**), where the animal is free to move, has been used in studies in the past (see Aubé et al., 2022, for a review; Battini et al., 2011; Destrez et al., 2018).

To be considered valid, a measure must be feasible, precise, stable over time, selective and sensitive (Knierim et al., 2021). Feasibility indicates the practical implementability of a measure and its ability to produce reliable results at acceptable cost and time-demand. Precision consists of intra-observer and inter-observer reliability. Intra-observer reliability is the degree of consensus when the same observer measures the same event under similar conditions (Knierim et al., 2021). Inter-observer reliability is the degree of consensus when several assessors measure the same event at the same time (Knierim et al., 2021). Precision can be assessed by a combination of reliability and agreement parameters (de Vet et al., 2006). Reliability parameters capture the capacity of a measure to distinguish two different individuals, despite measurement errors, and depend partly on the variability between study objects. Agreement parameters capture the measurement error and determine the closeness between results from repeated measurements (de Vet et al., 2006). A measure is stable over time if the results obtained at different times, with no major change in animal welfare on the farm, are similar. Stability over time can be assessed across a short-term window (usually called test–retest reliability, with measures repeated a few days apart, e.g. Rousing and Waiblinger, 2004) and across a long-term window (with measures repeated a few months apart, e.g. Battini et al., 2011; Winckler et al., 2007). A measure is considered selective if it allows to obtain true results of what it is assumed to assess, and not something else. Selectivity can be checked, for example, by convergent validity – which implies that two measures that are conceptually related are empirically associated – or by construct validity – by comparing the effects of conditions or treatments that reflect causal relationship (Waiblinger et al., 2006). The sensitivity of a welfare measure represents its capacity to discriminate between two different levels of welfare (Knierim et al., 2021).

The validity of ADF has been confirmed in terms of intra- and inter-observer reliability and convergent validity by comparing results against other measures reflecting HAR (Ebinghaus et al., 2018; Rousing and Waiblinger, 2004; Waiblinger et al., 2003; Windschnurer et al., 2009, 2008). ADF is stable over time when repeated after two months, but not at longer intervals (Battini et al., 2011; Winckler et al., 2007). The validity of ADP for dairy cows has not yet been established.

The aim of this study was to check key validity criteria of ADP as a measure of HAR in grazing dairy cows. We assessed ADP for inter-observer reliability and stability over short-term interval (test–retest) and long-term interval (stability over the season) and compared these performances against ADF. We also assessed the selectivity of ADP by convergent validity with ADF.

2. Materials and methods

2.1. Ethics statement

The experiment took place from April to November 2021 in the 'Le Pin-au-Haras' experimental farm (France, 48.448 N, 0.098E, DOI: 10.15454/1.5483257052131956E12) run by the French National Research Institute for Agriculture, Food and Environment (**INRAE**). The experiment was carried out according to the French guidelines for animal care and use and was approved by the institutional animal care and use committee (APAFIS agreement #20846–2019052810487566 v2). All procedures were performed by appropriately-trained personnel.

2.2. Animals, housing, and management

This study was part of a larger study on feeding management for dairy cows in pasture-based systems. A herd of 144 cows (48 Holstein, 48 Normande, 48 Jersey) housed indoors during winter were turned out to pasture starting March 18th, 2021 after calving. The cows were then at pasture all day long most of the time, except around milking twice a day. The cows were moved indoors for milking and, once milked, stayed for roughly 1 h in a pen where they had access to an automatic concentrate feeder. If the weather conditions were detrimental to the pasture (sudden or excessive rainfall or, conversely, severe drought characterized by the vegetation cover taking on a straw-like appearance), the cows were kept indoors at night, or in some cases all day. The cows were milked between 07h00 and 08h30 and between 16h00 and 17h30. All 144 cows grazed on the same pasture plots, which measured approximately 10 ha, resulting in an instantaneous stocking density of 14 cows/ha. A given plot was grazed for 7–12 days. The cows were dried off 90 days prior to expected parturition if their body condition scored below 2 (on a 0-to-5 scale, where 0 = emaciated and 5 = extremely fat), or were dried off when they produced less than 5 kg milk/d for seven consecutive days or no later than 60 days prior to parturition. Any cows that were dried off during the grazing season were removed from the grazing herd and thus excluded from the present study.

We selected 48 cows out of the total 144-cow herd: 16 Holstein, 16 Jersey and 16 Normande. Among each breed, 8 cows were primiparous and 8 were multiparous. This resulted in six possible combinations of breed × parity with 8 cows per combination. The 8 cows were randomly chosen within each combination. Over the course of the experiment, a total of five cows had to be excluded: four cows that were dried off (three cows due to insufficient milk production and one due to severe lameness), and one cow that died. Each time we excluded a cow, we replaced it with a cow of the same breed × parity combination. This resulted in a total of 53 cows used over the grazing season.

At the beginning of the experiment (April), cows were (mean ± SD) 50 ± 19.4 days in milk, 3.0 ± 0.9 years old (range 2.1–5.5 years), had 1.7 ± 0.8 parity, weighed 478 ± 114 kg, and had a body condition score of 3.0 ± 0.6. During the experimental period (from April to November), the cows produced 14.9 ± 6.5 kg milk/day. Weather and pasture conditions (days in the pasture plot, pasture access modalities) for each visit (i.e. each data collection session during the grazing season) are reported in supplementary material (Table S1).

2.3. Measurements and data collection

First, before starting the experiments, two persons (Person 1 and 2) were trained by two other persons, who have already followed the Welfare Quality training program and used to train persons to the WQ protocol, in a farm with dairy cows to assess ADF. Second, we checked the ability of Person 1 and 2 to estimate distances between themselves and fixed objects. Third, we assessed inter-observer reliability (i.e. between Person 1 and Person 2) for both ADF and ADP. Fourth, we assessed test–retest reliability and stability of ADP and ADF over the season when measured by the same person (Person 1). Fifth, we assessed the relationship between ADF and ADP. The two persons were both brown-haired women who were initially unfamiliar to these cows. Person 1 was 169 cm tall and Person 2 was 175 cm tall. During the tests, the two persons wore dark boots and were dressed in standard dark clothing (dark green overalls with black jackets).

2.3.1. Distance to objects

The ability of the Persons 1 and 2 to estimate distance was checked by asking them to estimate distances between themselves and fixed objects (wooden pins). We used 30 distances taken at random between 0 and 5 m, with a 0.10 m precision. A third person placed 30 wooden pins on the floor so that the distances between pins corresponded to the 30 selected distances. Persons 1 and 2 were then asked to estimate the distance between themselves and the pins to nearest 10 cm, and then five minutes later, the two persons repeated the exercise for the same 30 test distances, in the same order as in the first sequence.

2.3.2. Avoidance distance tests

For all avoidance tests described below, the person that performed the test (testing the cows and recording the result) is called “experimenter” and the person that observed the test (recording the result from a distance away) is called “observer”, as used in Windschnurer et al. (2009).

To assess ADF, the cows were headlocked at the feed-barrier indoors. When cows left the milking parlour in the morning, they were directed toward a pen as usual, and once all the cows had been milked, they were headlocked at the feed barrier indoor (as usually done for different handling procedures on the cows, e.g. blood sampling, oral bolus administrations, etc.). ADF was then assessed by the experimenter following the procedure set out in the Welfare Quality Consortium (2009) on the 48 selected cows (which placed themselves randomly at the feeding rack). The experimenter stood 2 m in front of the cow’s head, then approached the cow at a speed of one step per second with one arm at 45° in front of the body (palm facing downwards). The experimenter stopped walking when the cow accepted contact with the hand or when the cow showed a sign of avoidance, as defined in Welfare Quality Consortium (2009), i.e. the cow “moves back, turns the head to the side, or pulls back the head trying to get out of the feeding gate; head shaking can also be found”. The avoidance distance corresponds to the distance, estimated by sight, between the experimenter’s hand and the cow’s muzzle at first sign of avoidance (with a precision of 10 cm; from 0 cm – the cow can be touched – to 200 cm – the cow moves at the beginning of the test).

ADP was assessed when the cows came back to pasture after the morning milking. All the cows to be tested were clearly marked with numbers from 1 to 48, in order to be easily findable on pasture. The person(s) stood still in the pasture plot as the cows arrived and remained still until the cows scattered for grazing (around 5 min). The experimenter moved to be 5 m in front of a cow among the 48 selected cows that was nearest to the experimenter and that was standing (grazing or idling). The distance of 5 m to begin the test was chosen based on a preliminary experiment in which all the cows could be approached at this distance without any avoidance reactions. Then, the experimenter approached the cow in the same way as for ADF (one step per second, arm at 45°, palm facing downwards) and stopped at the first sign of avoidance. ADP varied from 0 cm– cow can be touched – to 500 cm – cow shows avoidance movement from the beginning of the test. During the ADF and ADP tests, the observer stood 4 m from the experimenter on her side.

2.3.3. Experimental design to check the avoidance test validity criteria

The validity of the two avoidance tests were checked based on three criteria: inter-observer reliability, test–retest reliability, and stability over the season (Table 1).

Inter-observer reliability was checked in ADF and ADP by comparing the results of the two persons, whatever their roles, i.e. as experimenter or as observer. This was done on two consecutive days. On the first day, Person 1 had the role of the experimenter for the first 24 cows (cows numbered 1–24) while Person 2 was the observer. The roles were reversed for the remaining 24 cows (cows numbered 25–48): Person 1 being the observer and Person 2 being the experimenter. On the second day, the 24 cows (cows numbered 1–24) that had been tested by Person 1 the first day, were tested by Person 2, and vice versa for the remaining 24 cows (cows numbered 25–48) (Table 1). During the experiment, the next cow to be tested was chosen by identifying the one, among the 48 selected ones, that was closest to the two persons at pasture, and the one who was next at the feeding rack. Therefore, the order in which the cows were tested for both ADF and ADP was random, i.e. the next cows tested could be as well as a cow that was assigned to be tested by Person 1 or by Person 2. For each avoidance test, both the experimenter and the observer estimated the avoidance distance, giving 96 measures of ADF and 96 measures of ADP for each person over the two days (giving 48 measures estimated while in the experimenter role and 48 while in the observer role for each person).

Test–retest reliability (**TRT**) was checked by Person 1 performing the ADF test twice, one day apart (October 4th and 5th 2021) (Table 1). Stability over the season (**STA**) was checked by Person 1 performing ADF and ADP tests on 7 visits at 5 week-intervals over the grazing season (Table 1).

2.4. Statistical analyses

Statistical analyses were performed using R software (v.4.1; R Core Team, 2021). Nonparametric tests were used due to the non-normality of the majority of the data.

2.4.1. Reliability and agreement between two series of measures

The reliability of results between two series of distances includes reliability between true distances to objects and the persons’ estimates, inter- and intra-observer reliability for distance-to object estimates, the inter-observer reliability and test–retest reliability of ADF and ADP at cow-level. Reliability was checked by calculating Spearman’s rank correlation coefficient (**RS**) and concordance correlation coefficient

Table 1
Experimental design and statistical parameters of agreement and reliability for the performance criteria of the measures (inter and intra-observer reliability, test–retest reliability, and stability over the season).

Measures	Criteria	Experimental design	Number of data	Agreement parameters	Reliability parameters
Distance estimation with fixed objects	Intra-observer reliability Inter-observer reliability Comparison between true distances and persons’ estimations	Person 1 and Person 2 estimated the 30 distances twice (i.e. two repetitions)	N = 30 distances estimations per person per repetition	·Systematic bias: median of the difference between the two series of measures ·LoA: 2.5% and 97.5% percentiles of difference between the two series of measures ·SDC: smallest change in the measure that can be detected given the measurement error ·Percentage of agreement between the two series of measures according to different tolerance windows (from 10 to 70 cm)	·RS: Spearman’s rank correlation ·CCC: concordance correlation coefficient
ADF and ADP (Visit 1)	Inter-observer reliability	Day 1: Cows 1–24: Obs1 - Exp2 and Cows 25–48: Obs2 - Exp1 Day 2: Cows 1–24: Obs2 - Exp1 and Cows 25–48: Obs1 - Exp2	N = 47 cows per person for ADF N = 42 cows per person for ADP		
ADF and ADP (Visit 6)	Test–retest reliability	Day 1: Cows 1–48: Exp 1 Day 2: Cows 1–48: Exp 1	N = 46 cows per day		
ADF and ADP (Visit 1 to Visit 7)	Stability over the season	Cows 1–48: Exp 1 Repeated seven times, at a 5-week interval	N = 42 cows per visit	Percentage of agreement between the seven series of measures according to different tolerance windows (from 10 to 70 cm)	·CCC: concordance correlation coefficient ·KW: Kendall’s coefficient of concordance

ADF, avoidance distance at the feeding rack; ADP, avoidance distance at pasture; Exp1: Person 1 being the experimenter, Exp2: Person 2 being the experimenter; Obs1: Person 1 being the observer; Obs2: Person 2 being the observer.

(**CCC**), using U-statistics with the 'cccrm' package (Carrasco et al., 2013; Carrasco and Martinez, 2022) (Table 1). RS was considered low when below 0.40, moderate from 0.41 to 0.70, high from 0.71 to 0.90, and very high at 0.91 or above (Martin and Bateson, 1993). CCC was considered low when below 0.20, fair from 0.21 to 0.40, moderate from 0.41 to 0.60, high from 0.61 to 0.80, and very high at 0.81 or above (Landis and Koch, 1977).

The agreement between two series of distances was checked by calculating i) systematic bias as the median of the difference between the two series of measures, ii) limits of agreement (**LoA**) as the 2.5% and 97.5% percentiles of this difference (where a higher amplitude of LoA indicates lower agreement), and iii) percentages of agreement with different distance tolerance windows (between 10 and 70 cm) using the 'irr' package (Gamer et al., 2019). We also calculated the smallest detectable change (**SDC**) using the 'AgRee' package (Feng, 2020) as an indicator of agreement. SDC gives information on the smallest change in the measure that can be detected (Friedrich et al., 2019) (Table 1).

2.4.2. Stability over the season

To assess the stability of the measures over the seven visits at cow-level, we calculated CCC and Kendall's concordance coefficient (**KW**) using the 'irr' package (Gamer et al., 2019). KW was considered low when below 0.4, moderate from 0.41 to 0.70, high from 0.71 to 0.90, and very high at 0.91 or above (Martin and Bateson, 1993) (Table 1).

Moreover, in order to check whether ADF and ADP differed between visits, we investigated the effect of the visit on avoidance distance using two mixed linear models, one for ADF and another for ADP, using the 'lme4' package (Bates et al., 2015), with visit as fixed factor and individual cows as random factor. Residuals were not normally distributed for ADF, so the raw data were transformed into log(ADF + 10) values.

2.4.3. Relation between ADP and ADF

The relation between ADP and ADF, considering all visits during the season, was tested using a linear mixed model with ADP as the variable to be explained, ADF as fixed factor, and individual cows as random factor.

3. Results

3.1. Reliability and agreement for estimates of distances to objects

Between observers' estimates and true distances. In tests assessing distances to fixed objects, the two persons achieved very high reliability with true distances, with coefficients (RS and CCC) above 0.91 (Table 2). The amplitude of LoA was around 150 cm for Person 1 and 175 cm for Person 2. The bias (i.e. the median of the difference between true distances and observers' estimates) was 5 cm for Person 1 and 10 cm for Person 2. SDC was 74 cm for Person 1 and 125 cm for Person 2. Within a tolerance window of 10 cm, Person 1 achieved 60% of agreement with true distance, while Person 2 achieved 43% of agreement (Fig. 1). Within a tolerance window of 30 cm, Person 1 achieved 77% of agreement with true distance, while Person 2 reached 60%. Within a tolerance window of 60 cm, Person 1 achieved 90% of agreement with true distance, while Person 2 achieved 73% of agreement.

Intra-observer reliability. In tests assessing distances to fixed objects, the two persons obtained very high intra-observer reliability, with coefficients (RS and CCC) above 0.96 (Table 2). The bias between the two estimations (i.e. the median of the difference between the two estimates of each person) was 5 cm for Person 1 and 0 cm for Person 2. SDC was around 65 cm for both persons. Within a tolerance window of 10 cm, Person 1 achieved 53% agreement and Person 2 achieved 43% (Fig. 1). The two observers achieved 83% agreement within a tolerance window of 40 cm and 90% agreement within a tolerance window of 50 cm.

Inter-observer reliability. In tests assessing distances to objects, the two persons obtained very high inter-observer reliability, with coefficients (RS and CCC) above 0.93 (Table 2). The amplitude of LoA was around 160 cm, the bias was 10 cm, and SDC was around 100 cm. The two persons achieved 50% agreement within a 20 cm tolerance window, 77% agreement within a 50 cm tolerance window, and 87% within a 70 cm tolerance window (Fig. 1).

3.2. Inter-observer reliability for ADF and ADP

For ADF, the final sample was composed of 48 cows on the first day and 46 cows on the second day (due to cows being excluded from the herd; see 2.2), leading to a total of 94 data points per person. For ADP, out of the initial sample of 48 cows, 43 cows were tested the first day and 41 cows the second day (due to some cows being excluded from the herd, see 2.2; and some cows that could not be tested as they were lying down on the pasture), leading to a total of 84 data points per person.

For ADF, inter-observer reliability was fair considering CCC (0.32) and high considering RS (0.78). For ADP, inter-observer reliability was very high considering CCC (0.83) and high considering RS (0.87) (Table 3). The amplitude of LoA between Person 1′s estimates and Person 2′s estimates was 119.3 cm for ADF and 249.3 cm for ADP (Table 3). The bias (i.e. the median of the difference between Person 1′s estimates and Person 2′s estimates) was 10 cm for both ADF and ADP. For ADF, the two persons agreed 69% of the time within a 10 cm tolerance window and achieved more than 75% of agreement within a 20 cm tolerance window (Fig. 2). For ADP, the two persons agreed 50% of the time in a 40 cm tolerance window and achieved 75% of agreement within a 70 cm tolerance window (Fig. 2).

Regardless of which person had the role of observer vs. experimenter, the observer overestimated ADP compared to the experimenter, with a systematic bias of 25 cm, whereas there was no systematic bias for ADF. This systematic bias could be due to the fact that the observer is situated further away from the cow and from the experimenter for ADP compared to ADF.

Table 2
Intra- and inter observer reliability and agreement; reliability and agreement between persons' estimations and true distances for estimated distances to fixed objects.

	Intra-observer Person 1	Intra-observer Person 2	Between observer 1′s estimation and true distance	Between observer 2′s estimation and true distance	Inter-observer
	N = 30	N = 30	N = 30	N = 30	N = 30
Reliability					
Spearman rank correlation	0.97 (P < 0.001)	0.97 (P < 0.001)	0.96 (P < 0.001)	0.96 (P < 0.001)	0.96 (P < 0.001)
CCC [95%CI]	0.97 [0.96–0.98]	0.98 [0.96–0.99]	0.96 [0.94–0.98]	0.92 [0.87–0.95]	0.94 [0.91–0.96]
Agreement					
LoA (cm)	(−62.8) to 63.8	(−48.3) to 71.0	(−50) to 98.3	(−25) to 152.8	(−32.8) to 131
Bias (cm)	5	0	5	10	10
SDC [95%CI] (cm)	65.6 [52.4–87.7]	66.9 [53.5–89.5]	74.2 [59.3–99.2]	125.5 [100.5–168.2]	103.6 [82.8–138.5]

CCC, concordance correlation coefficient; LoA, limits of agreement; SDC, smallest detectable change

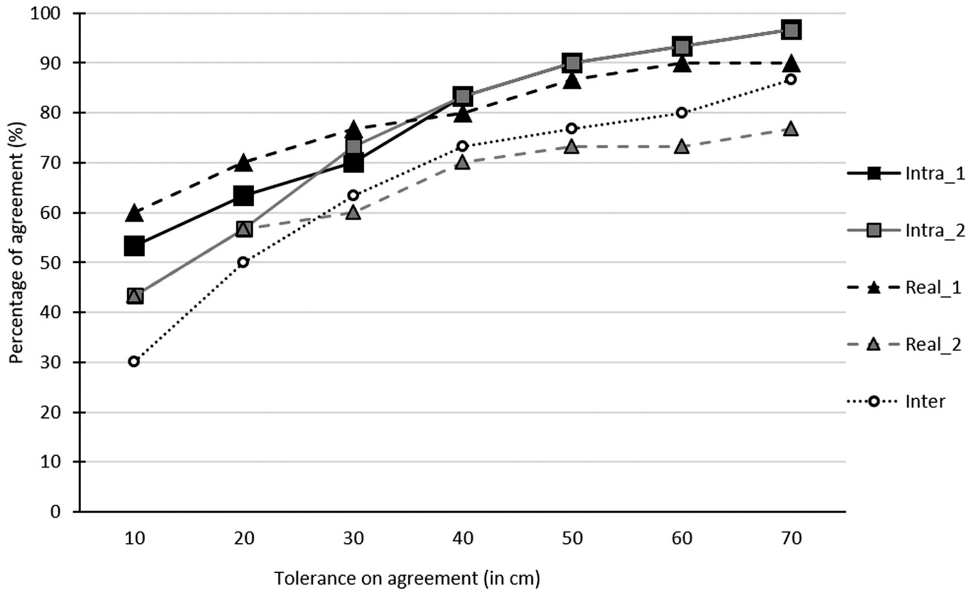

Fig. 1. Percentage of agreement between the first and the second estimates of the distance between fixed objects for Person 1 and Person 2 (Intra_1 and Intra_2), between the first estimate of each person and the true distance (True_1 and True_2), and between the first distance estimates of the two persons (Inter).

Table 3
Reliability and agreement between observers or repetitions for both avoidance distance at the feeding rack (**ADF**) and avoidance distance at pasture (**ADP**).

	Inter-observer	Between observer and experimenter	Test–retest	Stability over the season
ADF				
Number of data	N = 94	N = 94	N = 46	N = 42
Reliability				
Spearman's rank correlation	0.78 (P < 0.001)	0.64 (P < 0.001)	0.67 (P < 0.001)	*not calculable*
KW [95%CI]				0.58
CCC U-statistic [95%CI]	0.32 [0.17–0.46]	0.26 [0.11–0.41]	0.69 [0.49–0.82]	0.62 [0.52–0.70]
Agreement				
LoA (cm)	(−35.8) to 83.5	(−60) to 50	(−47.5) to 48.8	*not calculable*
Bias (cm)	10	0	0	*not calculable*
SDC [95%CI]	54.5 [47.7–63.6]		206.2 [171.4–259.0]	*not calculable*
ADP				
Number of data	N = 84	N = 84	N = 46	N = 42
Reliability				
Spearman's rank correlation	0.87 (P < 0.001)	0.88 (P < 0.001)	0.50 (P < 0.001)	*not calculable*
KW				0.57
CCC U-statistic [95%CI]	0.83 [0.78–0.87]	0.83 [0.78–0.87]	0.42 [0.28–0.55]	0.48 [0.39–0.55]
Agreement				
LoA (cm)	(−119.3) to 130	(- 79.9) to 148.5	(−187.5) to 208.8	*not calculable*
Bias (cm)	-10	25	25	*not calculable*
SDC [95%CI]	127.1 [110.5–149.7]		206.2 [171.4–259.0]	*not calculable*

CI, confidence interval; KW, Kendall's coefficient of concordance; CCC, concordance correlation coefficient; LoA, limits of agreement; SDC, smallest detectable change; WQ, Welfare Quality

3.3. Test–retest reliability of ADF and ADP

Over the two days, 46 cows were tested twice for both ADF and ADP.

For ADF, test–retest reliability was high considering CCC (0.69) and moderate considering RS (0.67). For ADP, test–retest reliability was moderate considering CCC (0.42) and RS (0.50) (Table 3). Amplitude of LoA was around 100 cm for ADF and 400 cm for ADP (Table 3). No systematic bias between the two days was found for ADF whereas there was a systematic bias of 25 cm for ADP (higher distances on the first day of test than the second day of test). For ADF, the percentage of agreement reached 70% for a 10 cm tolerance window and 90% for a 50 cm tolerance window (Fig. 2). For ADP, the percentage of agreement reached 50% for a 50 cm tolerance window and was still below 70% for a 70 cm tolerance window (Fig. 2).

3.4. Stability over the season of ADF and ADP

Among the 53 cows actually included in the present study over the season, the final sample consisted of 42 cows that had been tested for ADF at all visits (the other cows having at least one missing datapoint over the 7 visits), 42 cows that had been tested for ADP at all visits, and 41 cows that had been tested for both ADP and ADF at all visits.

The average (i.e. mean) distance across cows varied between visits, from 24.6 to 29.8 cm for ADF and from 139 to 182 cm for ADP (see supplementary Table S3). The minimal and maximal values (min–max) observed on individual cows over the season were 0–120 cm for ADF and 0–450 cm for ADP.

For ADF, stability over the season was high considering CCC (0.62) and moderate considering KW (0.52). For ADP, stability over the season was moderate considering CCC (0.48) and KW (0.57) (Table 3). For ADF, the percentage of agreement over the 7 visits reached more than 75% in a 50 cm tolerance window and reached 90% in a 70 cm tolerance

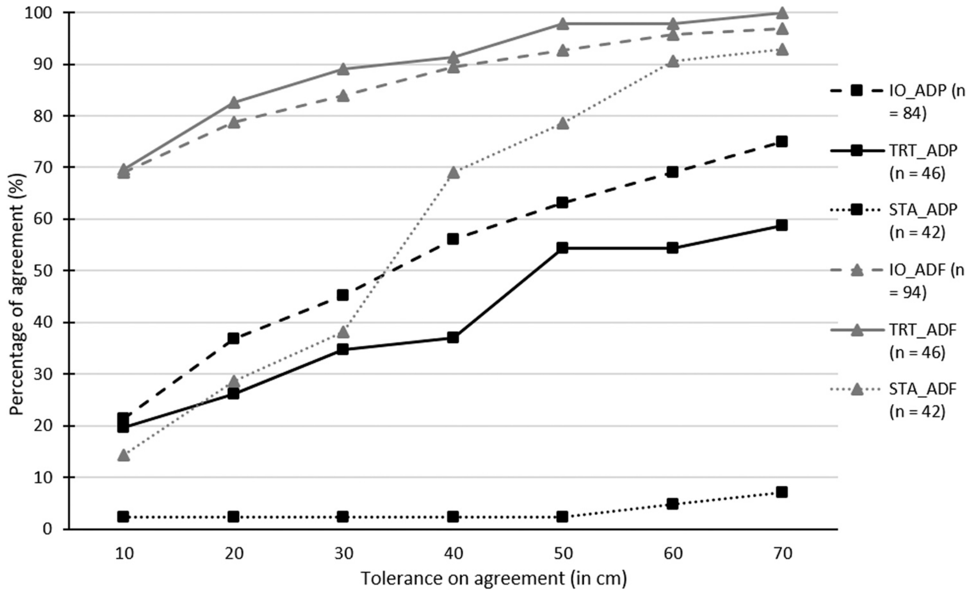

Fig. 2. Percentage of agreement between Persons 1 and 2 (IO); between the two days for test–retest (TRT) and between the seven visits (STA), with tolerance windows from 10 to 70 cm, for both avoidance distance at the feeding rack (ADF; grey lines) and avoidance distance at pasture (AD; black lines).

window, whereas it stayed below 10% for ADP even in the 70 cm tolerance window (Fig. 2). There was a significant effect of visit on ADP but not on ADF: compared to Visit 1 (intercept [95%CI] at Visit 1: 182 cm [152.1–211.6 cm]), ADP was significantly lower on Visit 2 (t value = −2.13) and Visit 7 (t value = −2.01) (Fig. 3).

Pairwise correlations between visits indicated mostly moderate correlation for ADF (with 19 RS between 0.4 and 0.7 and two RS ≤ 0.4) and moderate correlation for ADP (with 16 RS between 0.4 and 0.7, 4 RS below 0.4 and 1 RS above 0.7) (Table 4).

3.5. Relation between ADF and ADP

The relation between ADP and ADF was not statistically significant (ADF estimate [95%CI] = 0.14 [−0.31 to 0.60]; t value = 0.60; intercept = 160.4; n = 53).

4. Discussion

This study, involving a single herd at an experimental unit, provides first elements of validation of avoidance distance at pasture (**ADP**) as a

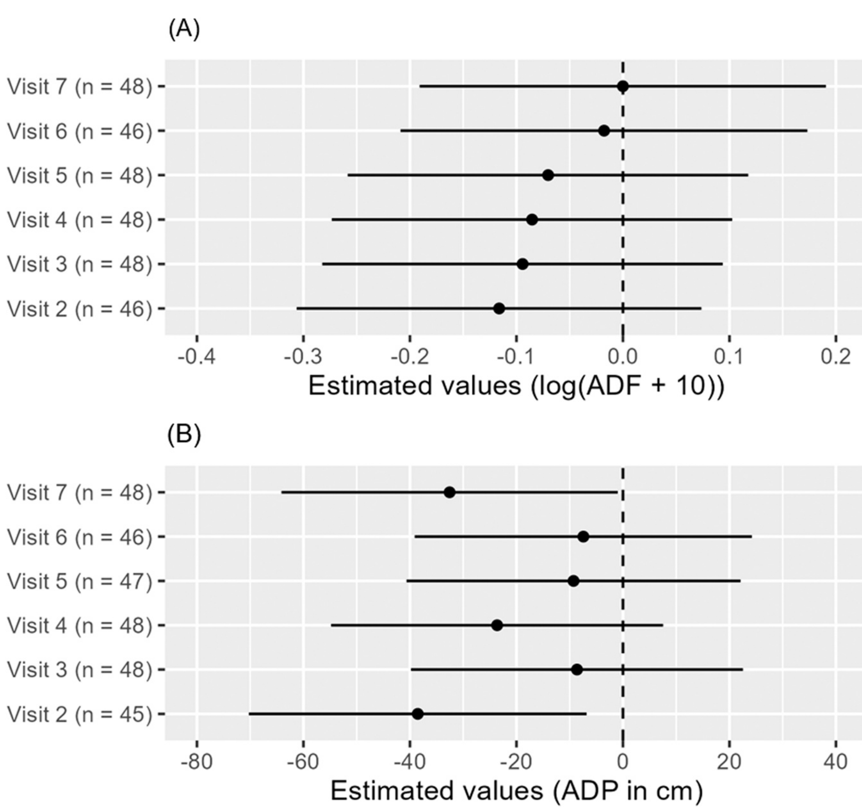

Fig. 3. (A) Avoidance distance at the feeding rack (ADF) (log-transformed data: (ADF + 10)) and (B) Avoidance distance at pasture (ADP) in cm for each visit relative to the first visit in the season, with 95% confidence intervals (CI). Relative values for each visit can be interpreted as an additive effect compared with the reference level (Visit 1). The intercept for ADP, which is the estimated mean of ADP at the first visit (n = 48), is equal to 182 cm (CI = [152.1–211.6]). The intercept for ADF, which is the estimated mean ADF at the first visit (n = 48) when back-transformed, is equal to 22 cm (CI = [16.3–28.9]).

Table 4
Relationship (Spearman's rank correlation coefficients*) between the different visits for avoidance distance at the feeding rack (ADF; gray background) and avoidance distance at pasture (ADP; in bold) assessed by Person 1 as experimenter.

		ADF Visit 1	ADF Visit 2	ADF Visit 3	ADF Visit 4	ADF Visit 5	ADF Visit 6	ADF Visit 7
ADP	Visit 1		0.50 n = 46	0.52 n = 47	0.51 n = 47	0.45 n = 47	0.35 n = 44	0.47 n = 43
ADP	Visit 2	**0.58 n = 45**		0.50 n = 46	0.66 n = 46	0.49 n = 46	0.50 n = 43	0.49 n = 42
ADP	Visit 3	**0.41 n = 47**	**0.57 n = 45**		0.52 n = 48	0.41 n = 48	0.48 n = 45	0.40 n = 44
ADP	Visit 4	**0.49 n = 47**	**0.39 n = 45**	**0.44 n = 48**		0.57 n = 48	0.60 n = 45	0.54 n = 44
ADP	Visit 5	**0.58 n = 46**	**0.39 n = 44**	**0.50 n = 47**	**0.60 n = 47**		0.54 n = 45	0.55 n = 44
ADP	Visit 6	**0.52 n = 44**	**0.36 n = 43**	**0.32 n = 45**	**0.53 n = 45**	**0.63 n = 45**		0.43 n = 45
ADP	Visit 7	**0.51 n = 43**	**0.73 n = 42**	**0.45 n = 44**	**0.53 n = 44**	**0.50 n = 44**	**0.42 n = 45**	

*Spearman's rank correlation was considered low when below 0.4, moderate from 0.41 to 0.70, high from 0.71 to 0.90, and very high at 0.91 or above (Martin and Bateson, 1993).

measure of human–animal relationship in dairy cows. When tested at cow-level, ADP had an acceptable inter-observer reliability but was not stable with time - neither for test–retest reliability nor for stability over the grazing season – and was not related to avoidance distance at the feeding rack (**ADF**).

The ability of the two persons to estimate distances to objects at 0–5 m was very high. Indeed, they both obtained very high reliability between their estimates and true distances. In addition, their estimates did not vary substantially across repetitions (high intra-observer reliability), and the inter-observer reliability was also good. The two persons were thus considered able to competently measure ADF and ADP, as the distance to be estimated is between 0 and 5 m. This is the first time, to our knowledge, that the ability of observers to estimate distances is tested before performing the avoidance distances tests.

The two observers obtained good reliability results for ADF. Spearman's rank correlation coefficient (**RS**) for inter-observer reliability was high (0.78) and similar to values reported in the literature (e.g. RS = 0.69 in Windschnurer et al., 2008; RS = 0.79 in Ebinghaus et al., 2016). ADF was also stable at a short-term interval, with high test–retest reliability (RS = 0.67). ADF presented moderate stability over the season considering Kendall's coefficient of concordance (**KW**) (0.58) and high stability considering concordance correlation coefficient (**CCC**) (0.62). These results are fairly consistent with previous literature reporting high test–retest reliability (Forkman and Keeling, 2009) but low stability over long-term intervals (Battini et al., 2011).

The two persons also achieved good inter-observer reliability results for ADP. Inter-observer RS (0.87) was even higher than for ADF. This may come from the distribution of data, as ADP produced data from 0 to 450 cm whereas ADF only varied from 0 to 120 cm, and a shorter range of variations can result in lower reliability coefficients (de Vet et al., 2006). However, the agreement between observers appeared to be less satisfactory for ADP than for ADF considering absolute values (amplitude of LoA $\simeq$ 250 cm vs. 120 cm, SDC 127 vs. 54 cm). This almost certainly comes from the distribution of data too. When the values for amplitude of LoA and SDC are expressed as percentage of the maximum distance (5 m for ADP and 2 m for ADF), they are similar between ADP and ADF (25–27% for amplitude of LoA and 50–60% for SDC).

Inter-observer reliability and agreement were lower for ADP than for distance to objects. When measuring an avoidance distance, the experimenter (performing the test) and the observer had to estimate the distance between the experimenter's hand and the cow's muzzle at the exact moment when the cow initiated the first avoidance movement. The cows often changed position immediately after the first avoidance movement, and so the persons had to remember the position of the muzzle at the time of withdrawal, which is harder to do than estimating the distance between fixed objects. Cows showed less ample movement at the feeding rack than at pasture (personal observations), being freer to move at pasture, and so the distance estimation is probably easier for ADF than for ADP. There may also be some uncertainty about the moment when the cow moves backward, i.e. the moment when the experimenter judged that the cow had withdrawn. When estimating avoidance distances, the observer was placed further away than the experimenter and had a different viewing angle of the cow, whereas when estimating distances to objects, the two persons were both similarly positioned in relation to the objects, which could explain the lower inter-observer reliability for ADP than for distance estimation with fixed objects.

The test–retest reliability of ADP was moderate (RS = 0.50) and lower than for ADF (RS = 0.67). In addition, results showed that large cow-level variability in ADP from one day to the next. Indeed, the amplitude of LoA between two successive days was almost 4 m, meaning that although the ranking of cows within the herd is moderately stable, their actual distance can vary a lot. In our opinion, these results suggest that the reliability and agreement parameters used were not satisfactory between the two days to validate the test–retest reliability of ADP. There was a systematic bias of 25 cm between Day 1 and Day 2 for ADP (higher distances on the first day of test than the second day of test). This bias is not observed for ADF and could be due to the larger observation distances for ADP. This bias was nevertheless lower than the margin of error of Person 1, and so this result should be taken with caution. Low bias for ADP and the absence of bias for ADF between the two days suggest that there was no habituation of the cows to the tests from one day to another. The large cow-level variability in ADP from one day to the next could be explained by the fact that the cows were not tested in the same order from one day to the next. Thus, the variable length of time spent by a cow in presence of the experimenter at pasture before getting tested may affect cow responses, either due to habituation to the experimenter's presence or because motivation to feed changes with time during the day.

ADP was moderately stable over the grazing season. The moderate reliability coefficients (KW = 0.57 and CCC = 0.48) obtained when the observations were repeated every 5 weeks from April to November implies that the ranking of cows according to their ADP was moderately stable from one visit to another. In comparison, ADF appeared to be a little bit more stable over the season than ADP, with KW = 0.58 and CCC = 0.62. However, less than 10% of the cows presented an ADP within a 70-cm window across all seven visits, suggesting high intra-individual variability between visits. This was not the case for ADF (with more than 75% of the cows presenting a distance within a 40-cm window across all seven visits). At herd level, the mean ADP of cows differed by a maximum of 40 cm between two visits. This difference is included in the margin of error of Person 1, and consequently does not allow to firmly conclude on a true difference between visits at herd level. We therefore find that ADP does not have stability over the season at individual level but seems stable at herd level. This could be explained probably by the fact that individual cows varied in their responses from one visit to another. Indeed, ADP of some cows decreased from one visit to another while it increased for others cows, resulting in an average ADP quite similar between visits at herd level. From a welfare point of view, it is possible that ADP reflects the reactivity of the cows to humans at one time point but does not reflect the whole HAR. Even if the measure seems stable at herd level as it is not stable at individual level further studies in several herds seem necessary to investigate the stability of this measure.

The variation of ADP over the season is not a surprising result. Dodzi and Muchenje (2011) found significant differences in ADP scores between repetitions carried out several months apart during four contrasted seasons. Note that this variation in ADP score differed according to cow genotype (either Jersey, Friesland, or crossbred). Dodzi and Muchenje (2011) suggested that fear of humans could be influenced by season (and thus probably by temperature), but further studies are needed to explore these variations. Here, the individual-level variation in ADP from one visit to another could probably be explained by several factors, both human and environmental factors. Battini et al. (2011) found an increased avoidance distance at the end of the pasture season at herd level due to less human exposure during grazing season, but the effect of this human factor on individual-level variation was not assessed and could be different. Environmental factors could also have an impact as pasture environment varied between visits—weather conditions, vehicles passing, animals present or not in adjacent plots, quantity and quality of grass, etc.—and some cows may have been more sensitive than others to these environmental changes. Feeding motivation can be a confounding factor (Waiblinger et al., 2006). Animals can be less interested and/or pay less attention to human presence if they are more attracted or motivated to feed. It is possible that the feeding motivation of the cows, in our study, was more important just after the cows came back to pasture after milking and gradually decreased as the cows grazed. Therefore, reactivity of the cows may have been influenced by the time spent grazing before the ADP test, and as the order of the tested cows was random this could partly explain the instability of ADP at individual level. The hour of the test was not recorded in the present study but this must be taken into account in future studies on ADP. The nutritional needs of cows also vary along the season depending on their physiological state, making them more or less sensitive to changes in grass quantity and/or quality. We suggest that the test should at least be standardized, e.g. for comparison between a set of farms, the grazing environment has to be largely similar between farms (same season, same amount of grass, cows at the same physiological stage, etc.).

We found no significant relationship between ADP and ADF at individual cow level. This is the first time, to our knowledge, that the link between ADF and ADP has been investigated at individual level in dairy cows. Dodzi and Muchenje (2011) explored the relation between ADP scores (on a 5-point scale from 1, where the cow avoided the experimenter at a distance > 2 m, to 5 where the cow let the experimenter touch it) and six measures that are thought to reflect HAR (exit speed, pen score, pen behaviour score, platform score, and the occurrence of kicking and stepping during milking). The results showed that ADP score was only related (correlation coefficient > 0.4) to two (exit speed and stepping at milking) out of the six tested measures (Dodzi and Muchenje, 2011).

Here, the lack of relation between ADF and ADP could be due to the different reactions of cows depending on the location where the tests were done (at pasture or at the feeding rack), as management change could have an impact on avoidance distance (Battini et al., 2011). Interactions with humans differ when the cows are at the feeding rack (where there are many interactions, such as food distribution, handling for blood tests, reproduction monitoring, etc.) and at pasture (where interactions essentially revolve around herding the cows together to travel from pasture toward the milking parlour). Cows may thus perceive humans differently depending on their location. Being approached at pasture was an unusual experience for these cows (except in the context of herding), and so they may have reacted more to the novelty of the situation than to the experimenter per se (Breuer et al., 2003). Finally, as suggested by Armbrecht et al. (2019), cows have lots of space at pasture and the ground is non-slippery – which is not always the case indoors – and so they can better express withdrawal behaviour. Despite this result, ADP remains a promising measure for assessing HAR in grazing cows, as a previous study (Dodzi and Muchenje, 2011) found relations between ADP scores and two other measures reflecting HAR.

Our results suggest that ADP is not adequately stable over the season at individual level but seems quite stable at herd level, which should be assessed by further studies to continue the validation process of ADP measure. Further studies should investigate the different factors that influence ADP and assess ADP on several herds in order to evaluate stability over the season at herd level.

Although this study failed to show that ADP is selective in reflecting HAR through the convergent validity process (relation with ADF), further investigation is warranted. Construct validation could be used to investigate the selectivity of ADP, i.e. HAR could be experimentally manipulated (e.g. measuring ADP variation between cows that had positive contacts with humans versus cows that had negative contacts, as in Lensink et al., 2000).

Further studies are required to continue exploring ADP validity criteria, especially feasibility in commercial farm conditions, and sensitivity. Feasibility was not checked here, as the experimental context was different to real-world practice. Indeed, the experimenter had to find the 48 selected cows included in the test sample out of 144 cows of the herd, at pasture, which takes time (30–45 min for ADF, and 1–1h30 for ADP), whereas in real practice, cows would be chosen randomly within the herd. In our opinion, the best time of the day to perform ADP is when cows return to pasture after the morning milking as this always happened in daylight (unlike after the afternoon milking) and, when the cows arrive on pasture after milking, they usually all graze for a period of time that allow the ADP tests to be performed. Sensitivity was not investigated here as the experiment took place on only one herd where all cows had received similar human contacts before. Further studies should be carried out to compare cows submitted to varying human contacts (e.g. very few contacts vs. positive contacts) or to compare herds from several commercial farms as for ADF (Windschnurer et al., 2009).

5. Conclusion

In conclusion, ADP has similar inter-observer reliability to that of ADF. ADP is not stable in the short term (contrary to ADF) nor in the long term (like ADF). ADP is thus not yet validated as a measure of HAR, at least at individual level. Further studies should investigate if ADP can nevertheless be used to distinguish herds.

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

The authors thank the staff of the INRAE experimental unit 'Le Pin' (DOI: 10.15454/1.5483257052131956E12) for their technical assistance and for animal care. The authors also thank Frédéric Launay, head of the experimental unit, and Luc Delaby, scientific advisor, for affording us access to the experimental unit and to the experimental cows.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.applanim.2023.105999.

References

Armbrecht, L., Lambertz, C., Albers, D., Gauly, M., 2019. Assessment of welfare indicators in dairy farms offering pasture at differing levels. Animal 13, 2336–2347. https://doi.org/10.1017/S1751731119000570.

Aubé, L., Mialon, M.M., Mollaret, E., Mounier, L., Veissier, I., de Boyer des Roches, A., 2022. Review: Assessment of dairy cow welfare at pasture: measures available, gaps to address, and pathways to development of ad-hoc protocols. Animal 16, 100597. https://doi.org/10.1016/j.animal.2022.100597.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Battini, M., Andreoli, E., Barbieri, S., Mattiello, S., 2011. Long-term stability of avoidance distance tests for on-farm assessment of dairy cow relationship to humans in Alpine traditional husbandry systems. Appl. Anim. Behav. Sci. 135, 267–270.

Breuer, K., Hemsworth, P.H., Coleman, G.J., 2003. The effect of positive or negative handling on the behavioural and physiological responses of nonlactating heifers. Appl. Anim. Behav. Sci. 84, 3–22. https://doi.org/10.1016/S0168-1591(03)00146-1.

Carrasco, J.L., Martinez, J.P., 2022. cccrm: Concordance Correlation Coefficient for Repeated (and Non-Repeated) Measures. R package version 2.1.0. ⟨https://CRAN.R-project.org/package=cccrm⟩.

Carrasco, J.L., Phillips, B.R., Puig-Martinez, J., King, T.S., Chinchilli, V.M., 2013. Estimation of the concordance correlation coefficient for repeated measures using SAS and R. Comput. Methods Prog. Biomed. 109, 293–304.

Coleman, G.J., Hemsworth, P.H., Hay, M., 1998. Predicting stockperson behaviour towards pigs from attitudinal and job-related variables and empathy. Appl. Anim. Behav. Sci. 58, 63–75. https://doi.org/10.1016/S0168-1591(96)01168-9.

Destrez, A., Haslin, E., Elluin, G., Gaillard, C., Hostiou, N., Dasse, F., Zanella, C., Boivin, X., 2018. Evaluation of beef herd responses to unfamiliar humans and potential influencing factors: an exploratory survey on French farms. Livest. Sci. 212, 7–13. https://doi.org/10.1016/j.livsci.2018.03.011.

Dodzi, M.S., Muchenje, V., 2011. Avoidance-related behavioural variables and their relationship to milk yield in pasture-based dairy cows. Appl. Anim. Behav. Sci. 133, 11–17. https://doi.org/10.1016/j.applanim.2011.04.014.

Ebinghaus, A., Ivemeyer, S., Knierim, U., 2018. Human and farm influences on dairy cows′ responsiveness towards humans – a cross-sectional study. PLoS One 13, e0209817. https://doi.org/10.1371/journal.pone.0209817.

Ebinghaus, A., Ivemeyer, S., Rupp, J., Knierim, U., 2016. Identification and development of measures suitable as potential breeding traits regarding dairy cows' reactivity towards humans. Appl. Anim. Behav. Sci. 185, 30–38. https://doi.org/10.1016/j.applanim.2016.09.010.

Feng, D., 2020. agRee: Various Methods for Measuring Agreement. R package version 0.5–3. ⟨https://CRAN.R-project.org/package=agRee⟩.

Forkman, B., Keeling, L.J., 2009. Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves (Welfare Quality Report No. 11). School of City and Regional Planning, Cardiff University, Cardiff, UK.

Friedrich, L., Krieter, J., Kemper, N., Czycholl, I., 2019. Test−retest reliability of the 'Welfare Quality® animal welfare assessment protocol for sows and piglets'. Part 1. Assessment of the welfare principle of 'Appropriate behavior. Animals 9, 398. https://doi.org/10.3390/ani9070398.

Gamer, M., Lemon, J., Singh, I.F.P., 2019. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. ⟨https://CRAN.R-project.org/package=irr⟩.

Knierim, U., Winckler, C., Mounier, L., Veissier, I., 2021. Developing effective welfare measures for cattle. In: Understanding the Behaviour and Improving the Welfare of Dairy Cattle. Burleigh Dodds. https://doi.org/10.19103/AS.2020.0084.05.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Lensink, B.J., Boivin, X., Pradel, P., Le Neindre, P., Veissier, I., 2000. Reducing veal calves' reactivity to people by providing additional human contact1. J. Anim. Sci. 78, 1213–1218. https://doi.org/10.2527/2000.7851213x.

Martin, P., Bateson, P., 1993. Measuring Behavior: An Introductory Guide, second ed. Cambridge University Press, Cambridge, UK.

Napolitano, F., Bragaglio, A., Sabia, E., Serrapica, F., Braghieri, A., Rosa, G.D., 2020. The human−animal relationship in dairy animals. J. Dairy Res. 87, 47–52. https://doi.org/10.1017/S0022029920000606.

P.H. Hemsworth, Human–animal interactions in livestock production Appl. Anim. Behav. Sci., Int. Soc. Appl. Ethol. Spec. Issue.: A Sel. Pap. ISAE Int. Congr. 1999–2001 81 2003 185 198 doi: 10.1016/S0168-1591(02)00280-0.

R Core Team, 2021. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rousing, T., Waiblinger, S., 2004. Evaluation of on-farm methods for testing the human–animal relationship in dairy herds with cubicle loose housing systems—test–retest and inter-observer reliability and consistency to familiarity of test person. Appl. Anim. Behav. Sci. 85, 215–231. https://doi.org/10.1016/j.applanim.2003.09.014.

de Vet, H.C.W., Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59, 1033–1039. https://doi.org/10.1016/j.jclinepi.2005.10.015.

Waiblinger, S., Menke, C., Coleman, G., 2002. The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. Appl. Anim. Behav. Sci. 79, 195–219. https://doi.org/10.1016/S0168-1591(02)00155-7.

Waiblinger, S., Menke, C., Fölsch, D.W., 2003. Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. Appl. Anim. Behav. Sci. 84, 23–39. https://doi.org/10.1016/S0168-1591(03)00148-5.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.-V., Janczak, A.M., Visser, E.K., Jones, R. B., 2006. Assessing the human–animal relationship in farmed species: a critical review. Appl. Anim. Behav. Sci. 101, 185–242. https://doi.org/10.1016/j.applanim.2006.02.001.

Welfare Quality Consortium, 2009. Welfare Quality® assessment protocol for cattle. Welfare Quality Consortium, Lelystad, The Netherlands.

Winckler, C., Brinkmann, J., Glatz, J., 2007. Long-term consistency of selected animal-related welfare parameters in dairy farms. Anim. Welf. 16, 197–199.

Windschnurer, I., Boivin, X., Waiblinger, S., 2009. Reliability of an avoidance distance test for the assessment of animals' responsiveness to humans and a preliminary investigation of its association with farmers' attitudes on bull fattening farms. Appl. Anim. Behav. Sci. 117, 117–127. https://doi.org/10.1016/j.applanim.2008.12.013.

Windschnurer, I., Schmied, C., Boivin, X., Waiblinger, S., 2008. Reliability and inter-test relationship of tests for on-farm assessment of dairy cows' relationship to humans. Appl. Anim. Behav. Sci. 114, 37–53. https://doi.org/10.1016/j.applanim.2008.01.017.