VALIDITY AND FEASIBILITY OF HUMAN-ANIMAL RELATIONSHIP TESTS FOR ON-FARM WELFARE ASSESSMENT IN DAIRY GOATS

Monica Battini[a], Sara Barbieri[a], Susanne Waiblinger[b], Silvana Mattiello[a]

[a]Università degli Studi di Milano, Dipartimento di Scienze Veterinarie e Sanità Pubblica, 20133 Milano, Italy monica.battini@unimi.it, sara.barbieri@unimi.it, silvana.mattiello@unimi.it

[b]University of Veterinary Medicine Vienna, Institute of Animal Husbandry and Animal Welfare, Department for Farm Animals and Veterinary Public Health, Veterinaerplatz 1, 1210 Vienna, Austria susanne.waiblinger@vetmeduni.ac.at

Corresponding author: Monica Battini, Università degli Studi di Milano, Dipartimento di Scienze Veterinarie e Sanità Pubblica, 20133 Milano, Italy; Telephone: +39 02 58318033; Fax: +39 02 58318030; Email: monica.battini@unimi.it

Abstract

This study aims at establishing suitable tests to measure the quality of the Human-Animal Relationship (HAR) in dairy goats for on-farm welfare assessment protocols by evaluating the predictive validity of different categories of HAR tests and their feasibility in on-farm condition. Twelve commercial dairy farms in Northern Italy were selected and classified as “good” (n=6) and “poor” (n=6) HAR on the basis of reports from a technical advisor. Some variables were tested to measure the HAR: sneezing – the number of alert sounds; voluntary approach (VA) test – in a stationary situation, the latency to the first contact by goat and the % of goats that entered in contact with the test person and within a 1.5 m radius around her were recorded, both continuously and at scan sample intervals; avoidance distance (AD) test – the avoidance distance from a moving and approaching person and the % of tested goats, of contacts with the test person followed by withdrawal or of acceptances of gentle stroke were recorded. The feasibility of each test was evaluated based on costs, time consumption, safety and training requirement. Sneezes were rarely and random expressed, so they do not seem suitable to be included in a protocol. Both Principal Component Analysis and One-way ANOVA confirmed the predictive validity of most of the variables, which were able to discriminate between “good” and “poor” HAR farms. Latency to first contact resulted valid (P=0.01) and a high feasibility was reported. The indicator was easy to be recorded and the VA test could be stopped immediately after the first contact, saving time. Variables from the AD test resulted valid (% tested goats:

P=0.006; AD mean: P=0.016; % contacts: P=0.006; % acceptance: P=0.003), although they were more time consuming or required a more specific training compared to latency to first contact. The correlation among variables seemed to support also a convergent validity of the tests used. The investigation pinpointed promising behavioural tests to be included into on-farm welfare assessment protocols in dairy goats. Taking into account species, test results and feasibility considerations, we suggest the inclusion of latency to the first contact with the test person into on-farm welfare assessment protocols. However, these results should be further tested in a larger number of farms of different dimensions, to overcome the limitations of this study due to the small sample size and to check the effect of farm size.

Keywords

Dairy goats, human-animal relationship, welfare assessment, avoidance distance, behavioural tests

1. Introduction

The Human-Animal Relationship (HAR) is commonly defined as the degree of relatedness or distance between animals and humans, i.e. their mutual perception (Waiblinger et al., 2006a). The perception and consequently the responses of animals towards humans may be influenced by different factors: the genetic and underlying personality traits (e.g. fearfulness/emotionality; Visser et al., 2001), that may play an important role in the reactivity that animals show when interacting with humans; the experience of positive handling in early life stages, that is reported to have durable effects in some species (e.g. goats; Lyons, 1989), but not in others (e.g. dairy cows; Boissy and Bouissou, 1988); the stockperson behaviour, that is considered one of the major factors able to influence the reaction of animals towards humans. In fact, the quality of HAR in farm animals may be influenced by the number, duration and nature of daily interactions with the stockpeople (Estep and Hetts, 1992; Hemsworth and Coleman, 2010). Many studies have confirmed the sequential relationship of the stockperson attitude, behaviour and the reaction of animals (e.g. Lensink et al., 2000; for review: Waiblinger et al., 2006a). A good HAR may help to reduce the perceived aversiveness of some procedures (Boivin et al., 2000; Lensink et al., 2001); on the contrary, a poor HAR may induce fear and distress that worsen the perception of farm practices and negatively affect both animal welfare (Rushen et al., 1999) and production (Hemsworth, 2003; Hemsworth and Barnett, 1991; Lensink et al., 2001; Lyons, 1989). Therefore, HAR is strictly linked to animal welfare, and should be taken into account in welfare assessment schemes. To this aim, behavioural tests for assessing the quality of HAR due to previous experiences are commonly adopted at farm level (Jackson and Hackett, 2007; Waiblinger et al., 2006a).

HAR tests used in on-farm welfare assessment mainly belong to two categories: tests measuring the reaction either to a stationary person or to a moving person (Waiblinger et al., 2006a). These tests resulted valid, feasible and reliable in several species (sheep: Napolitano et al., 2011; fattening bulls: Windschnurer et al., 2009a; buffalos: de Rosa et al., 2005; dairy cows: Waiblinger et al., 2002, 2003; Windschnurer et al., 2008).

A convergent validity (Kamphaus and Frick, 2005; Waiblinger et al., 2006a) was checked in these studies as different measures used to assess the quality of HAR were significantly correlated among them and/or with stockperson behaviour or attitude. Avoidance distance in dairy cows was validated further by showing its sensitivity to gentle human interactions (Windschnurer et al., 2009b). The above-cited tests were found highly consistent across time and a good repeatability was reported among different observers.

Feasibility can be evaluated considering time, financial and safety requirements. As to the first two concerns, tests used in the above-mentioned studies were relatively quick and simple to be adopted in on-farm situation. Furthermore, no specific instruments were required to perform the tests: this is a considerable advantage in economic terms. Concerning the safety of the observer, some studies reported limitations if the behaviour of the animals has to be collected from inside the pen. For example, sheep were observed to run towards the observer, due to their marked gregarious behaviour. This experience suggested that it was advisable collecting measures with animals gathered at the manger (Napolitano et al., 2011). Similar conclusions were drawn for the evaluation of HAR in fattening bulls (Windschnurer et al., 2009a): the avoidance distance at the feeding rack was preferable and safer than performing the test inside the pen. However, in dairy cows the validity of tests performed outside the pen seemed lower than the validity of tests conducted in the pen (Waiblinger et al., 2003).

Both stationary and moving person tests can be performed when animals are in their home pen (familiar situation) or in a test arena (unfamiliar situation). Tests conducted in an arena are not suitable for the inclusion into on-farm welfare assessment protocols for feasibility reasons (e.g. moving the animals, building the test arena; Rousing and Waiblinger, 2004). Moreover, in the arena the reaction of animals may be influenced by a novel situation (e.g. stress induced by constraint, forced movement, isolation from the familiar group; Waiblinger et al., 2006a).

As for goats, so far the majority of HAR studies have used tests in an arena and scarce information is available about validity and feasibility of behavioural tests performed in the home pen. In this species, stationary, moving and pursuing person tests were performed and different variables were collected to evaluate the HAR in a test arena: latency to proximity, time in proximity, latency to contact, time in contact, and approach-withdrawal (Lyons, 1989; Lyons and Price, 1987; Lyons et al., 1988). Furthermore, the latency to approach a stationary person was adopted by Jackson and Hackett (2007) in a test arena to estimate the positive effect of a gentle handling treatment, whereas Mattiello et al. (2010) tested goats in their home pen, successfully applying the avoidance distance test to a moving person developed for cattle by Waiblinger et al. (2002, 2003) and further improved by Windschnurer et al. (2008; 2009a, b). However, Muri et al. (2013) reported strong avoidance behaviour and goats flocking around when the observer tried to perform the avoidance distance test described by Mattiello et al. (2010).

Therefore, some tests are available for evaluating HAR in goats, but most of them were only used in an experimental setting and are not feasible in the context of an on-farm welfare protocol, as they either are time consuming, or a test arena is necessary, while for the avoidance distance test performed in the home pen feasibility results by Mattiello et al. (2010) and Muri et al. (2013) are contradictory.

Further development of less intrusive and time consuming, but still valid methods for HAR evaluation in goats would be useful for the inclusion of this welfare issue into an on-farm welfare assessment scheme for this species (Battini et al., 2014). Given the circumstances, this study aims at establishing suitable tests to measure the quality of the HAR in dairy goats for on-farm welfare assessment protocols. The study evaluated the predictive validity of different categories of HAR tests and their feasibility in on-farm condition.

2. Material and methods

2.1 Farms

The study was conducted in January 2013 in Lombardy region (Northern Italy).

A technical advisor, who regularly provided assistance to dairy goat farms, was asked to select 6 farms with the best HAR and 6 with the worst HAR. The evaluation of the technical advisor was based on his experience with dairy goats management and his familiarity with the farms. HAR had never been specifically assessed before in those farms, as this is not part of the veterinary official evaluation in Italy. The advisor used his expertise and the evaluation was based on his impression on the stockperson attitude and behaviour towards the animals. According to this judgment, farms were classified as having a “good” HAR (n=6) or a “poor” HAR (n=6). In all the farms, lactating goats were housed in one single pen, with no access to outdoor run. The mean size of the assessed pens was 73.00±59.39 adult lactating goats (min 12; max 201). Although the mean pen size in “good” farms (36.83±10.43; min 12; max 84) was smaller than in “poor” farms (96.83±26.26; min 51; max 201), no statistical differences in pen size were present between “good” and “poor” HAR farms and some “good” farms were larger than some “poor” farms and *vice versa*. Goats were of Saanen and Alpine breeds, the two more widespread cosmopolitan dairy breeds.

2.2 Attitudinal questionnaire

In order to provide greater support of the advisor classification of the farms into “poor” and “good” HAR, the persons in charge of the handling of the animals were asked to answer to a questionnaire (partly modified from a questionnaire already adopted for dairy cows; Waiblinger et al., 2002) regarding their attitude towards goats and handling of goats (Table 1). On all farms only one person per farm, the farmers themselves, were caring for the goats. Stockpeople attitudes were shown to be the most important predictors of stockperson behaviour (for review see Hemsworth and Coleman, 2010). The attitudinal questionnaire was thus used to get information on farmer attitudes and some estimation of their behaviour.

2.3 HAR tests

Two researchers (the interviewer and the test person) conducted the farm visits. The interviewer was the only person informed about the farm classification and she administered the questionnaire to the farmers. The test person was the only person to perform the HAR tests and she was completely blind to the farm classification. She conducted the tests while the interviewer was filling the questionnaire. The test person was completely unknown to the animals, as she had no previous contact with them before starting the execution of the tests. The test person was a young female researcher, who had previously received an appropriate training and she always wore an overall similar to that used by the farmers. In order to avoid disturbance and possible confounding effects, all tests were performed in absence of the farmer or of any other person working on the farm.

Three different tests were performed following this order: 1) sneezing, 2) voluntary approach (VA) test and 3) avoidance distance (AD) test.

Sneezing: we defined “sneeze” the loud, high pitched, short and abrupt alert sound that goats produce to warn their mates about an imminent danger detected (Miranda-de La Lama and Mattiello, 2010). This vocalization is made with the mouth closed while the goat forcefully expels a single blast of air primarily through the nostrils. So far, no scientific evidence is available to support the use of sneezing for assessing the quality of HAR in goats. This indicator was included because farmers suggested alarm calls as possible indicator (Battini et al., 2014). The test procedure was the following: the test person entered the barn and immediately started walking through the feeding corridor, at a distance of 80 cm from the feeding rack, at a speed of one step/sec (60 cm/step). The number of sneezes heard from outside the pen was recorded. Immediately after this, the test person opened the pen gate, entered the pen, stopped and recorded the number of sneezes heard inside the pen in the precise moment when she entered. At the end of this test, the test person left the pen.

VA test: immediately after the sneezing test, the test person approached the pen again, waiting at the gate for 30 sec before re-entering; then she entered and walked to a pre-determined spot inside the pen close to the wall, stopped with her back against the wall, facing the goats, for 30 sec. After this, she marked a semi-circumference of 1.5 m radius by piling up some straw with the feet and then started the stopwatch. During the test, the test person stood motionless (back to the wall) for 5 min, without staring into goats’ eyes. The following data was recorded: latency (sec) to the first contact, which is defined as when the first goat nuzzles or touches any part of the test person’s body, including clothes, boots, stopwatch, hair, etc. (if no goats got into contact within 5 min, the time was stopped at 300 s); total number of goats that got into contact with the test person during the 5-min test; number of goats that got into contact with the test person at 1 min-scan intervals; number of goats within a 1.5 m radius around the test person, at 1 min-scan. At the end of this test, the test person left the pen.

AD test: the test person re-entered the pen immediately after the conclusion of the previous test. The AD test was adapted from the AD test conducted by Mattiello et al. (2010), following the suggestions by these authors. Once in the pen, the test person stood in front of a goat at a distance of 200 cm, established a reciprocal visual contact with the animal, then started to move slowly towards the animal at a speed of one step/s, 60 cm/step and the arm lifted with an inclination of 45°, the hand palm directed downwards, without looking into the animal's eyes, but looking at the muzzle. When the animal showed the first avoidance reaction (moving backwards, turning or shaking its head), the test person recorded the distance between her hand and the muzzle of the animal, with a resolution of 10 cm. This distance was estimated by sight; the accuracy of the estimates had been previously assessed during the training period. If the animal could be touched by the test person but immediately withdrew, this was recorded as "contact"; if, after the contact, the animal accepted gently stroking of the head for at least 3 sec, this was recorded as "acceptance". After being tested, each goat was identified (by marking it or reading the ear tag), in order to perform the test on all the animals when possible and to avoid testing the same goat. This procedure was then repeated on other goats, trying to test the whole herd. If at a certain point no more goats could be visually contacted for more than 10 min, the test was interrupted and the number of goats that could be tested until then was recorded.

The feasibility of each test was evaluated based on costs (e.g., specific equipment) and on the report from the test person considering time consumption, safety and training requirement.

2.4 Data analysis

Table 2 reports the variables calculated from the HAR tests.

Non-parametric analysis of variance (Mann-Whitney U test; Siegel and Castellan, 1992) was carried out in order to test for differences between "good" HAR and "poor" HAR farms according to the classification of the technical advisor; the results of the questionnaire were not used to re-classify farms into the two classes.

Spearman correlation ranks were calculated and Principal Component Analysis (PCA) was performed in order to investigate the relationships among all variables, except for Sneezes_OUT and Sneezes_IN, which were almost never recorded.

One farm classified by the technical advisor as "good" HAR rendered unexpected results in all HAR tests. In this farm, bucks were inside the pen when the tests were performed. We identified this as a procedural error; therefore, this farm was retested without bucks after three weeks. The results presented below refer to the tests performed with no bucks inside the pen.

Scores obtained from the answers to the questionnaire submitted to farmers (Table 1) were analysed by PCA in order to check their agreement with the classification into “good” and “poor” HAR based on the reports from the technical advisor. Only questions that had been successfully used in previous studies on dairy cattle (e.g. Waiblinger et al., 2002; Waiblinger et al., 2006b) were included in this analysis (Questions 1, 4-7, 11). These questions reflect general beliefs about goats as well as beliefs, behavioural intentions and affective attitudes about handling of goats.

The study followed the national ethical guidelines and met all humane standards, as no manipulation was necessary and only non-invasive observations were carried out on the subjects in their home pen. Farmers subjected to the interview were informed on the use of data and personal details collected during the farm visits. All information remained completely anonymous and limited to the specific framework of the project.

3. Results

Questionnaire to the farmers

The first two PCs of the questionnaire to the farmers explained 52.43% of total cumulative variance (Fig. 1). Most of the variables describing positive attitudes and behaviour had a high positive loading on PC1, with Questions 11a-11d, dealing on how much farmers indicate that they like to interact with the goats, showing the highest loadings (Fig. 1b). PCA on the answers to the questionnaire administered to farmers showed a trend of farms with “poor” HAR to cluster on the left side of the first PC and of farms with “good” HAR to cluster on the right side (Fig. 1a). However, the separation is not well defined. One farm that had been classified by the technical advisor as “poor” clustered on the extreme of the right side on the basis of the answers to the questionnaire, whereas a farm classified as “good” tended to cluster with “poor” farms (Fig. 1a).

Validity

Means (± s.e.) of all the variables were calculated (Table 3). Sneezes _OUT and Sneezes_IN were recorded with low frequencies, only in two farms (2.5% of Sneezes_OUT in one farm and 0.99% of Sneezes_IN in another farm); therefore, they were excluded from further analysis. Statistical differences were found in six variables out of ten (Table 3). Latency and AD were significantly lower in “good” HAR farms, whereas

%_contact_VA, %_contact_AD, %_acceptance_AD and %_tested goats_AD were significantly lower in "poor" HAR farms (Table 3). Spearman correlations highlighted strong relationships among some variables (Table 4). No correlation was found between %_1.5_radius_VA and %_tot_contact_VA with AD test variables (AD, %_contact_AD, %_acceptance_AD). These relationships were also confirmed by PCA (Fig. 2), where Latency and AD present high negative loadings on PC1 (low Latency and low AD correspond to "good" HAR), and are negatively correlated with the other variables (whose high values correspond to "good" HAR), that have high positive loadings on PC1 (Fig. 2). Based on these loadings, "good" (on the right) and "poor" (on the left) HAR farms are well separated on PC1 (Fig. 2).

Feasibility

All the tests performed did not require specific equipment, so costs were limited for their collection. Sneezes_OUT and Sneezes_IN could be easily and quickly recorded in less than 60 seconds. Being recorded from outside, Sneezes _OUT was also safe for the observer and not stressful for the animals. A simple training was required in order to be able to recognise sneezing and distinguish it from other sounds. However, their occurrence was very low.

No problems were encountered for performing the VA test, but it necessarily took 300 seconds, therefore the time required was longer than for Sneezes. Latency was the easiest variable to record within VA test, whereas counting the number of animals that approached and got in contact with the test person required an accurate training and its feasibility in pens with a large number of animals still has to be verified. Furthermore, the feasibility of %_1.5_radius_VA was further reduced by the time required to mark the semi-circumference.

AD test was the more time consuming test. The average time required was 10.42±1.31 min (mean±s.e.; range: 3-17 min), mainly depending on the number of tested goats (min 6, max 33 goats), which ranged from 4.48% (in the largest farm) to 100.00% (in small farms). It was not always possible to test all the goats in the pen, as some animals never accepted the visual contact with the test person at the starting distance. No dangerous situation was reported while performing the AD test, although in some cases the presence of the test person in the pen induced a large number of goats to flock around.

4. Discussion

The research identified some potentially suitable HAR tests to be included into on-farm welfare assessment schemes for dairy goats, evaluating them according to predictive validity and feasibility criteria.

The classification suggested by the technical advisor into "good" and "poor" HAR did not completely match with the answers to the questionnaire submitted to farmers. Although on dairy cattle farms similar questionnaires did correlate with stockperson behaviour (Waiblinger et al., 2002), the predictive value of such questionnaires is limited as indicated by only moderate correlation coefficients. Thus for some farmers the questionnaire outcome may not completely reflect their actual behaviour, because (1) behaviour is influenced by other factors, such as perceived and actual control over the situation, and (2) perceived social norms may influence the actual behaviour and/or lead to dishonest answers to a questionnaire (Hemsworth and Coleman, 2010; Waiblinger et al., 2007).

In fact, in our investigation, the "poor" HAR farm that clustered as being the best of the "good" HAR farms had been described by the technical advisor as being managed by a farmer who had valiant theoretical ideas on animal management and welfare, but could not practically actualize them. This farm actually clustered with those with "poor" HAR with respect to the PCA on HAR tests, in agreement with the classification of the technical advisor.

Although the technical advisor based his classification on stockpeople attitudes and behaviour, we cannot exclude that his experience with the goats and thus goat behaviour on the farms unconsciously influenced his evaluation and one may argue that the validity of results would in this case be reduced. However, his classification of farms nevertheless is an independent evaluation of the HAR and thus associations with the HAR tests support their predictive validity (Waiblinger et al., 2006a). Further, the questionnaire supported the classification of the technical advisor according to attitudes at least for most of the farms.

Although farmers suggested the use of sneezes to reveal fear in response to humans perceived as a possible danger, our research could not support this hypothesis, probably due to the method used to record the indicator that could be not optimal. For example, the walking speed could be too slow to elicit any fear reaction.

All VA test variables were significantly correlated among themselves; however, only %_contact_VA and Latency statistically differed between "good" and "poor" HAR farms. Latency results (Table 3) are consistent with those found by Jackson and Hackett (2007) who reported that gently handled goats approached the experimenter more quickly compared to control goats during the latency test (228 *vs* 419 sec, respectively). Latency was the most feasible indicator: it required no specific training, as the test person can easily recognise when a goat enters in contact with him/her, and it can last a very short time (between 0 and 300 sec), because the test is stopped when the first contact occurs. Latency also appeared to be very safe for the observer, as animals had no negative reaction nor showed aggressive behaviours towards the test person. The test person should necessarily enter the pen, but being a stationary test, goats were not very agitated and the test could be easily performed in all farm situations.

All variables recorded during the AD test were significantly different between "good" and "poor" HAR farms. As expected, low AD values corresponded to high %_contact_AD, %_acceptance_AD and %_tested goats_AD, as highlighted by PCA, although the correlation of this last variable with AD was not significant. As suggested by Mattiello et al. (2010), the AD test included %_contact_AD and %_acceptance_AD, both of which seem to be valid indicators of HAR quality. These two levels can possibly increase the sensitivity of the test, as highlighted also for dairy cattle by Rousing and Waiblinger (2004), who added further levels to the AD test, adapting two behavioural tests developed by Krohn et al. (1999) and Waiblinger et al. (2003). As to feasibility, AD test can be time consuming if a large number of goats has to be tested. Furthermore, AD requires specific training to properly move in the pen, recognise the first avoidance reaction and assess the correct distance if goats withdraw, whereas both % contact_AD and % acceptance_AD could be performed quite easily, as no specific training is required to instruct the test person in recognising and discriminating contact and acceptance (although a general training is always required for a correct execution of the test). In our research, we also considered %_tested goats_AD, which helped to complete the evaluation: our results suggest that the number of animals accepting the visual contact (previously defined as the beginning of the AD test) with the test person during the AD test can be affected by the quality of HAR. The strong avoidance behaviour reported by Muri et al. (2013) might be related to "poor" HAR in the tested farms: recording the percentage of animals that can be tested may be a useful information to understand HAR quality. In fact, in some of our "poor" HAR farms, we encountered a situation similar to that described by Muri et al. (2013) and we could not perform the test on all animals. This was the case, for example, in the largest farm, with more than 200 goats in the pen, where less than 5% of the animals could be tested. The correlation of this variable with the other variables supports its convergent validity; however, we have to acknowledge the fact that goats in larger pens may behave differently than in small pens, and this may affect the number of animals that could be tested. Furthermore, a large number of goats running around can result in a danger situation for the observer inside the pen. The test person who performed the AD test did not report any unsafe situation, but the use of AD test should be carefully evaluated.

A further note is necessary regarding the presence of bucks in the pen with females. When we entered the pens, we observed that no females approached the test person if bucks were present. Males were generally the first to approach the test person, and they inhibited the approach behaviour of female goats. Therefore, the results of the tests could be compromised by the presence of bucks in the pen, lowering validity and feasibility.

Although variables from both AD test and VA test discriminated between farms with differently assessed HAR on PC1, and thus show not only a predictive, but also a convergent validity to assess the HAR on the farm, they provided slightly different information, as shown by the opposite distribution on PC2 of AD variables (%_acceptance_AD and %_contact_AD) and VA variables (%_tot_contact_VA and %_1.5_radius_VA). PC2 might thus reflect differences in motivation or movement of the animal (active approach in the VA variables, passive acceptance of human approach in the AD variables). As suggested by Welp et al. (2001), the use of stationary *vs* moving person tests, or *vice versa*, should be carefully considered. Stationary person tests are frequently used for on-farm welfare assessment: the test design is generally easy to perform and the tests are not too time-consuming (Waiblinger et al., 2006a). Stationary person tests seem more suitable for species that rarely interact with humans, although curiosity may be a stronger motivation to approach the person than fear of (Waiblinger et al., 2003; Marchant et al., 1998). On the contrary, species or individuals that have non-traumatic routinely contacts with humans (e.g. dairy cows) may ignore the stimulus person and the different motivations (fear, disinterest) are hard to distinguish without detailed behavioural observations (Waiblinger et al., 2003, 2006a). Accordingly, in dairy cows a test with a stationary person in the home pen was less valid (correlated less with milker’s behaviour) than avoidance distance (Waiblinger et al., 2003). On the other hand, moving person tests may feign situation that animals experience every day, as the presence of stockperson inside the pen is common in many species. These tests are considered feasible for on-farm assessment of HAR, especially for herd assessment (Waiblinger et al., 2006a). Some factors may influence the reaction during moving person tests, thus the test execution requires particular attention and previous training to perform it: for example, the speed of moving or body postures of the test person are some important influencing factors (Waiblinger et al., 2006a). A study conducted by Welp et al. (2001) in dairy cows confirms our results and reports that both stationary and moving person tests are highly correlated at herd level, indicating that they strengthen each other and they are both valid to assess the HAR. However, it should be carefully considered if tests at herd level can replace each other (Welp et al., 2001).

5. Conclusion

The present research is intended to be a first study towards the identification of potentially suitable methods for the evaluation of HAR, to be included into on-farm welfare assessment schemes for dairy goats. Our results are promising and some valid and feasible indicators for HAR evaluation can be suggested. However, these results should be further tested in a larger number of farms of different dimensions to overcome the limitations of this study due to the small sample size, and to check the effect of farm size.

The selection of the most appropriate behavioural test is strictly related to the target species. For goat, which is a curious species that rarely interacts with humans in the home pen, a voluntary approach test seems to be the most suitable option. Taking into account species, test results and feasibility considerations, we suggest the inclusion of latency to the first contact with the test person into on-farm welfare assessment protocols. However, in order to avoid eliciting strong fear reactions when entering the pen, the preliminary test procedure adopted in this study (to wait at the gate before entering and to select a pre-determined spot inside the pen) is necessary in order to standardise the test. Furthermore, we suggest performing behavioural tests when males are not inside the pen with females or to temporarily separate bucks from the group.

6. Acknowledgment

The Animal Welfare Indicators (AWIN) project (FP7-KBBE-2010-4) has received funding from the European Union Seventh Framework Programme for research, technological development and demonstration. We acknowledge the farmers for their collaboration and Giorgio Esposti (SATA – Servizio di

Assistenza Tecnica agli Allevatori, Lombardy Region, Italy) for his help with farm selection and classification.

7. References

Battini, M., Vieira, A., Barbieri, S., Ajuda, I., Stilwell, G., Mattiello, S., 2014. Animal-based indicators for on-farm welfare assessment for dairy goats: a review. J. Dairy Sci. 97, 6625-6648. http://dx.doi.org/10.3168/jds.2013-7493.

Boissy, A., Bouissou, M.F., 1988. Effects of early handling on heifers' subsequent reactivity to humans and to unfamiliar situations. Appl. Anim. Behav. Sci. 22, 259–273.

Boivin, X., Tournadre, H., Le Neindre, P., 2000. Hand feeding and gentling influence early weaned lamb's attachment responses to their stockperson. J. Anim. Sci. 78, 879–884.

de Rosa, G., Napolitano, F., Grasso, F., Pacelli, C., Bordi, A., 2005. On the development of a monitoring scheme of buffalo welfare at farm level. Italian J. Anim. Sci. 4, 115–125.

Estep, D.Q., Hetts, S., 1992. Interactions, relationships and bonds: the conceptual basis for scientist–animal relations, in: Davis, H., Balfour, A.D. (Eds.), The Inevitable Bond—Examining Scientist–Animal Interactions. CAB International, Cambridge, UK, pp. 6–26.

Hemsworth, P., 2003. Human–animal interaction in livestock production. Appl. Anim. Behav. Sci. 81, 185–198.

Hemsworth, P.H., Barnett, J.L., 1991. The effects of aversively handling pigs, either individually or in groups, on their behaviour, growth and corticosteroids. Appl. Anim. Behav. Sci. 30, 61-72.

Hemsworth, P.H., Coleman, G.J., 2010. Human–Livestock Interactions. The Stock-person and the Productivity and Welfare of Intensively Farmed Animals. 2nd ed., CAB International.

Jackson, K.M.A., Hackett, D., 2007. A note: The effects of human handling on heart girth, behaviour and milk quality in dairy goats. Appl. Anim. Behav. Sci. 108, 332–336.

Kamphaus, R.W., Frick, P.J., 2005. Clinical Assessment of Child and Adolescent Personality and Behavior. 2nd ed. Springer, New York, NY.

Krohn, C.C., Foldager, J., Mogensen, L., 1999. Long-term effect of colostrum feeding methods on behaviour in female dairy calves. Acta Agri. Scand. A. Anim. Sci. 49, 57–64.

Lensink, J., Boissy, A., Veissier, I., 2000. The relationship between farmers’ attitude and behaviour towards calves, and productivity of veal units. Annales de Zootechnie. 49, 313-327.

Lensink, J., Raussi, S., Boivin, X., Pyykkönen, M., Veissier, I., 2001. Reactions of calves to handling depend on housing conditions and previous experience with humans. Appl. Anim. Behav. Sci. 70, 187–199.

Lyons, D.M., 1989. Individual differences in temperament of dairy goats and the inhibition of milk ejection. Appl. Anim. Behav. Sci. 22, 269–282.

Lyons, D.M., Price, E.O., 1987. Relationships between heart rates and behaviour of goats in encounters with people. Appl. Anim. Behav. Sci. 18, 363–369.

Lyons, D.M., Price, E.O., Moberg, G.P., 1988. Individual differences in temperament of domestic goats: consistency and change. Anim. Behav. 36, 1323–1333.

Marchant, J.N., Burfoot, A., Corning, S., Broom, D.M., 1998. Human approach test—a test of fearfulness or investigatory behaviour?, in: Hemsworth, P.H., Spinka, M., Costal, L. (Eds.), Proceedings of the 31^{st} International Congress of the ISAE, Prague, Czech Republic, p. 182.

Mattiello, S., Battini, M., Andreoli, E., Minero, M., Barbieri, S., Canali, E., 2010. Technical note. Avoidance distance test in goats: A comparison with its application in cows. Small Rumin. Res. 91, 215–218. http://dx.doi.org/10.1016/j.smallrumres.2010.03.002.

Miranda-de la Lama, G.C., Mattiello, S., 2010. The importance of social behaviour for goat welfare in livestock farming. Small Rum. Res. 90, 1–10. http://dx.doi.org/10.1016/j.smallrumres.2010.01.006.

Muri, K., Stubsjøen, S.M., Valle, P.S., 2013. Development and testing of an on-farm welfare assessment protocol for dairy goats. Anim. Welf. 22, 385–400. http://dx.doi.org/10.7120/09627286.22.3.385.

Napolitano, F., de Rosa, G., Girolami, A., Scavone, M., Braghieri, A., 2011. Avoidance distance in sheep: test-retest reliability and relationship with stockmen attitude. Small Rumin. Res. 99, 81–86.

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.

Rushen, J., Taylor, A.A., de Passille, A.M., 1999. Domestic animals' fear of humans and its effect on their welfare. Appl. Anim. Behav. Sci. 65, 285–303. http://dx.doi.org/10.1016/S0168-1591(99)00089-1.

Siegel, S., Castellan Jr., N.J., 1992. Statistica Non-Parametrica. 2^{nd} ed. McGraw-Hill, Milan, Italy.

Visser, E.K., Van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. Appl. Anim. Behav. Sci. 74, 241–258.

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.

Waiblinger, S., Menke, C., Fölsch, D.W., 2003. Influences on the avoidance and approach behaviour of dairy cows towards humans in 35 farms. Appl. Anim. Behav. Sci. 84, 23–39.

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.

Waiblinger, S., Mülleder, C., Menke, C., Coleman, G., 2006b. How do farmers' attitudes impact on animal welfare? The relationship of attitudes to housing design and management on dairy cow farms, in: Arnat, M., Mariotti, V. (Eds.), Proceedings of The importance of attitudes, values, and economics to the welfare and conservation of animals, Barcelona, Spain, pp. 55-56.

Waiblinger, S., Spoolder, H., 2007. Quality of stockpersonship. In: Velarde, A., Geers, R. (Eds.), On farm monitoring of pig welfare. Wageningen Academic Publishers, Wageningen, pp. 159-166.

Welp, T., de Pasillé, A.M., Rybarczyk, P., Rushen, J., 2001. How valid are measures of cows' fear of people. Proceedings of the 35th International Congress of the ISAE, University of California, Davis, August 4–9, p. 162.

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.

Windschnurer I., Barth,K., Waiblinger, S., 2009a. Can stroking during milking decrease avoidance distances of cows towards humans? Animal Welfare 18, 507-513.

Windschnurer, I., Boivin, X., Waiblinger, S., 2009b. 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.

Figure captions

Fig. 1 – PCA results (PC1 vs PC2; 52.43% of cumulative explained variance) using variables derived from the answers given to the attitudinal questionnaire: a) loadings plot (variables' distribution); b) score plot (farms' distribution).

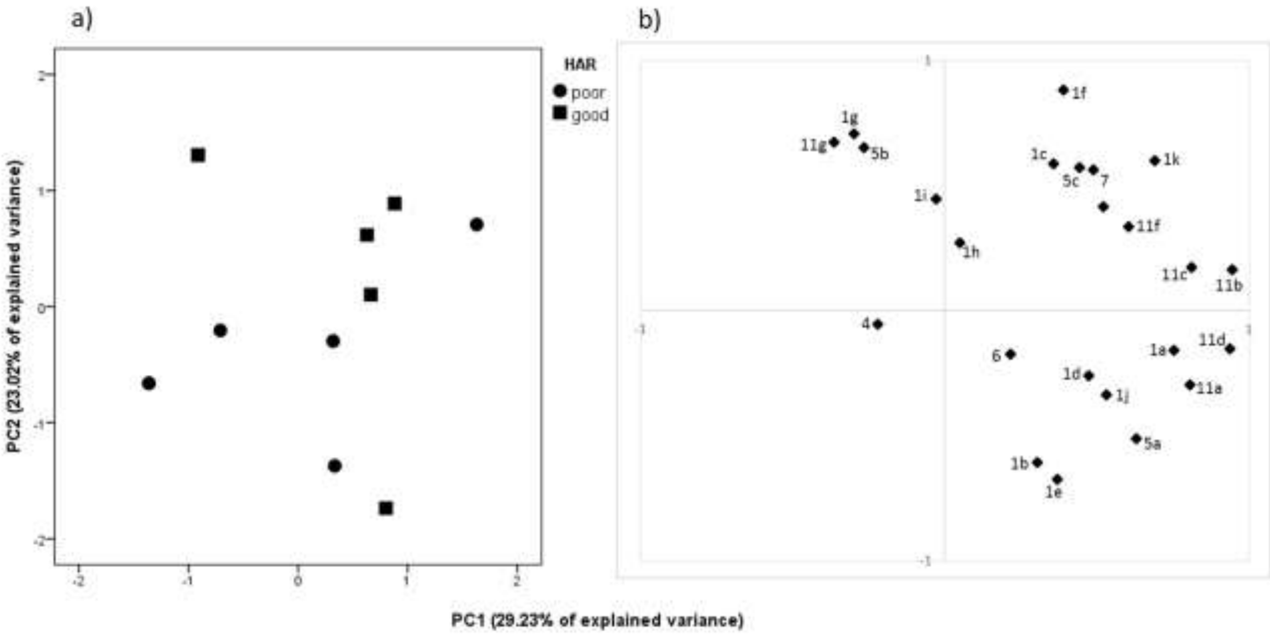

Fig. 2 – PCA results (PC1 vs PC2; 82.05% of cumulative explained variance) using variables from HAR tests on "good" and "poor" HAR farms: a) loadings plot (variables' distribution); b) score plot (farms' distribution).