Supplementary Material (Paper ID: 605)

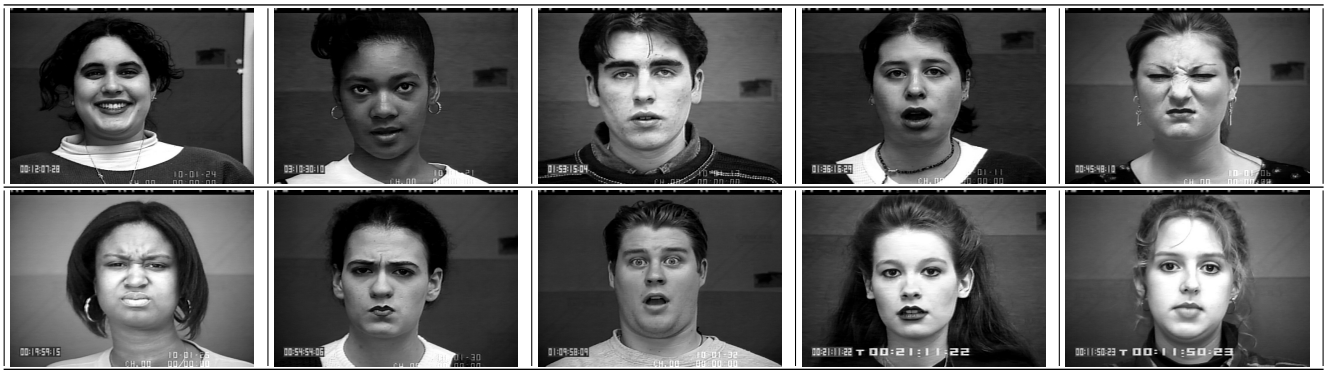
This supplementary material shows additional Figures that support our arguments in the main paper. The sections follow the same structure as the main paper.

(1) Introduction

- Main Paper lines: $\mathbf{142 - 146}$

a)

b)



c)

Figure 2: The differences between our FEEL database when compared with the facial-based approaches for emotion recognition: (a) FEEL database - images annotated with 26 feeling categories and the 3 common continuous dimensions *Valence, Arousal* and *Dominance*; (b) CK (Cohn-Kanade) database [21] - composed of frontal face images in gray scale generated in a highly controlled environment. Images are coded according to action units derived from FACS (Facial Action Coding System); (c) EMOTIONET database [3] - composed strictly of face images found using noun and verb relationships for the word *feeling* in WordNet. This is different from CK in that this has color images and the facial expressions are spontaneous. Unlike (b) & (c), FEEL includes the contexts, surroundings and all the nearby objects in the images, thereby enriching it with more information content. *For example:* We can see in (a,1) that the woman is in pain from some suffering and the other woman is trying to sympathize with her. Similarly we can see in (a,5) that an old man is completely engaged in his painting - these kind of information is absent in other datasets (b),(c). FEEL also includes images with faces not clearly visible. *For example:* In (a,2) & (a,3), the faces are obscured, however, their pose, attention, objects they hold and corresponding backgrounds provide information about their feeling.

(3.1) Image Annotation

- Main Paper lines: $\mathbf{266 - 269}$

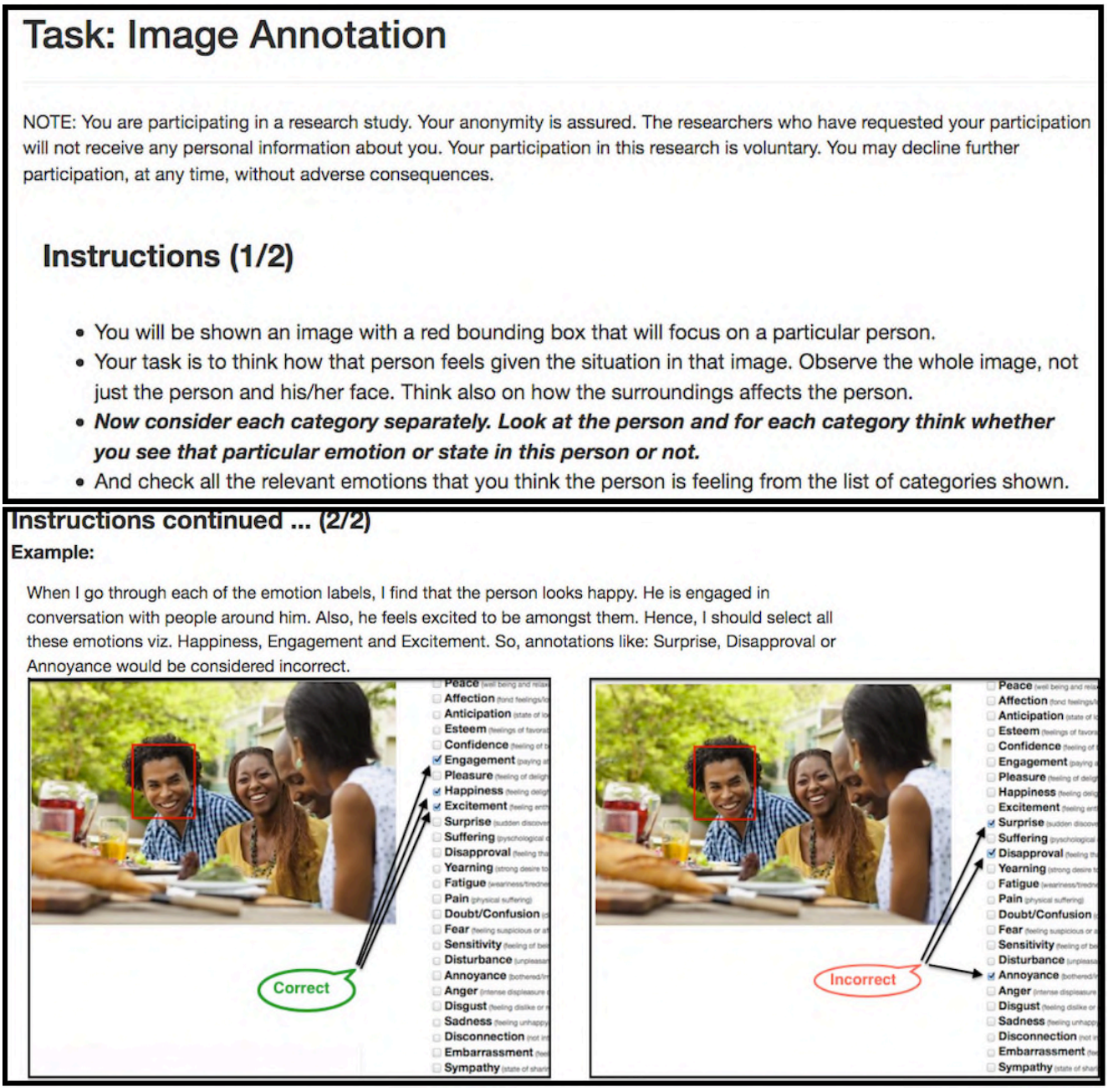


Figure 3: The instructions on how to attempt the categorical HITs are clearly explained in the interface. It also displays an example of correct and incorrect annotations, which serves as a guide to the workers for demonstrating on how to perform the actual task. We have also included these instructions and the sample example on every page of the actual HIT.

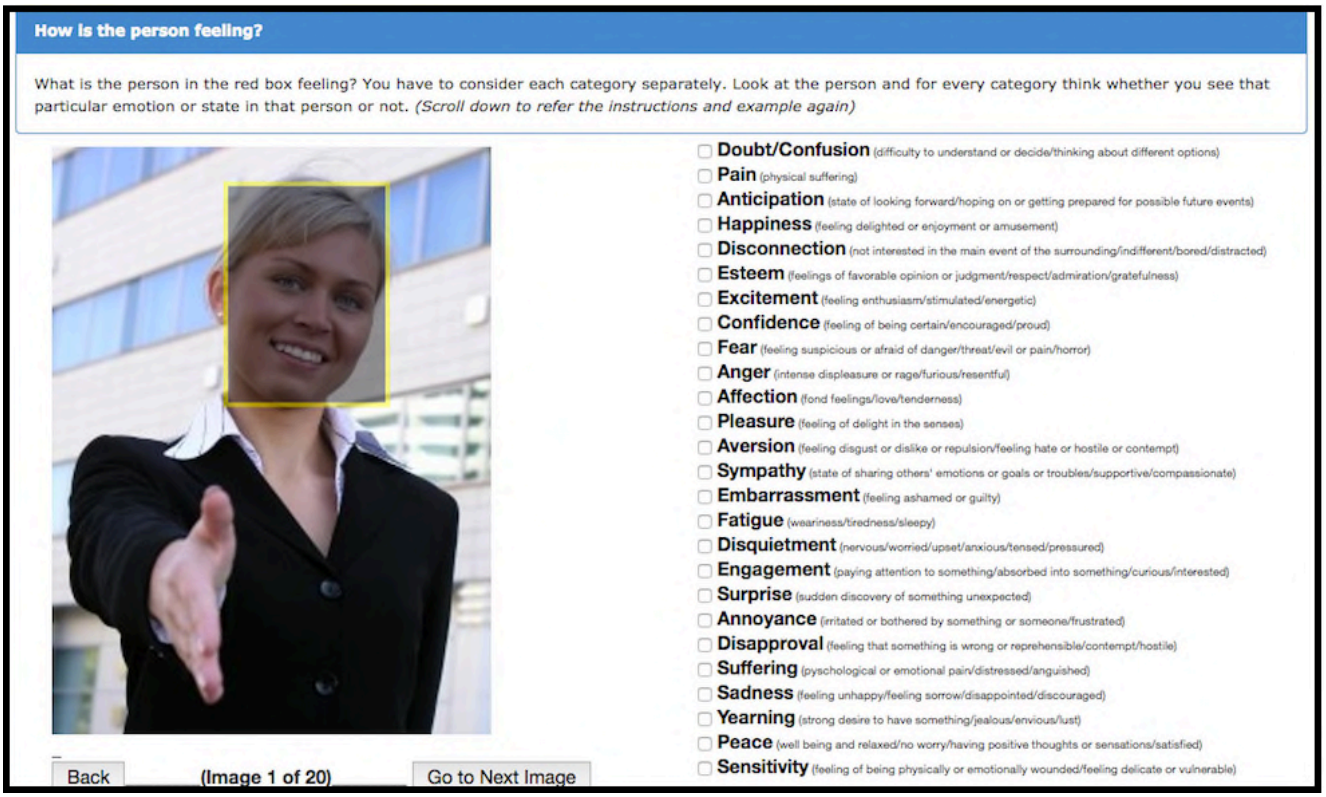


Figure 4: This is what the annotation task looks to the workers. The workers choose the feeling categories that are applicable to the person shown in the particular situation in the image.

(3.1) Image Annotation

- Main Paper lines: $\mathbf{305 - 308}$

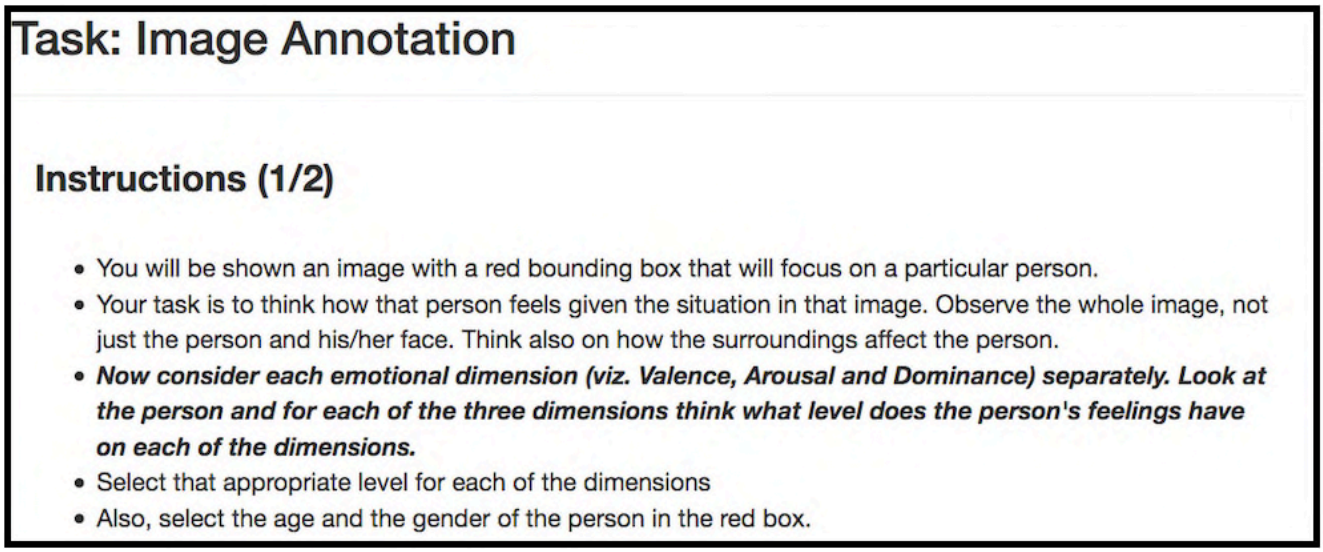


Figure 5: The instructions for the Continuous HIT task are clearly mentioned in the interface.

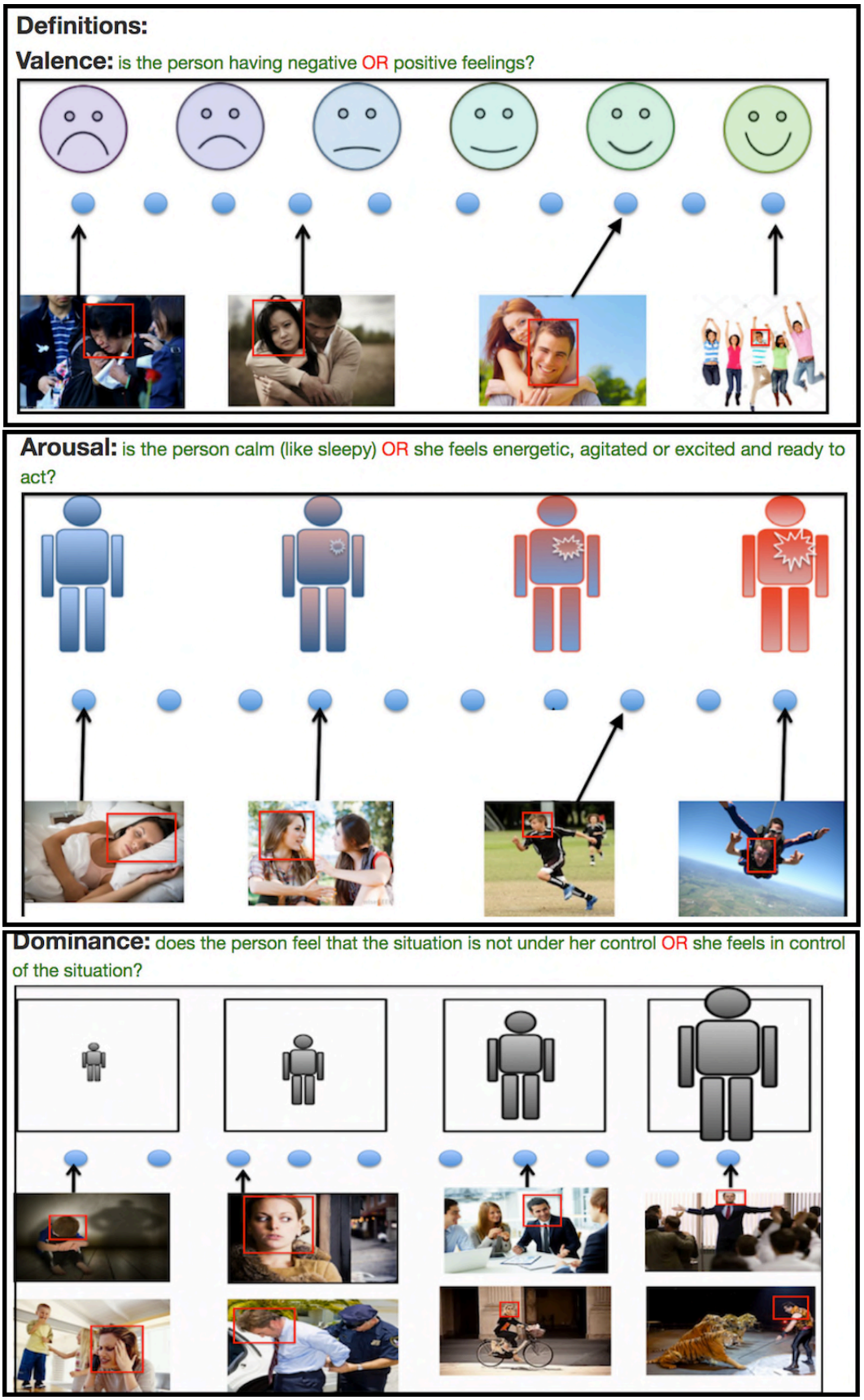
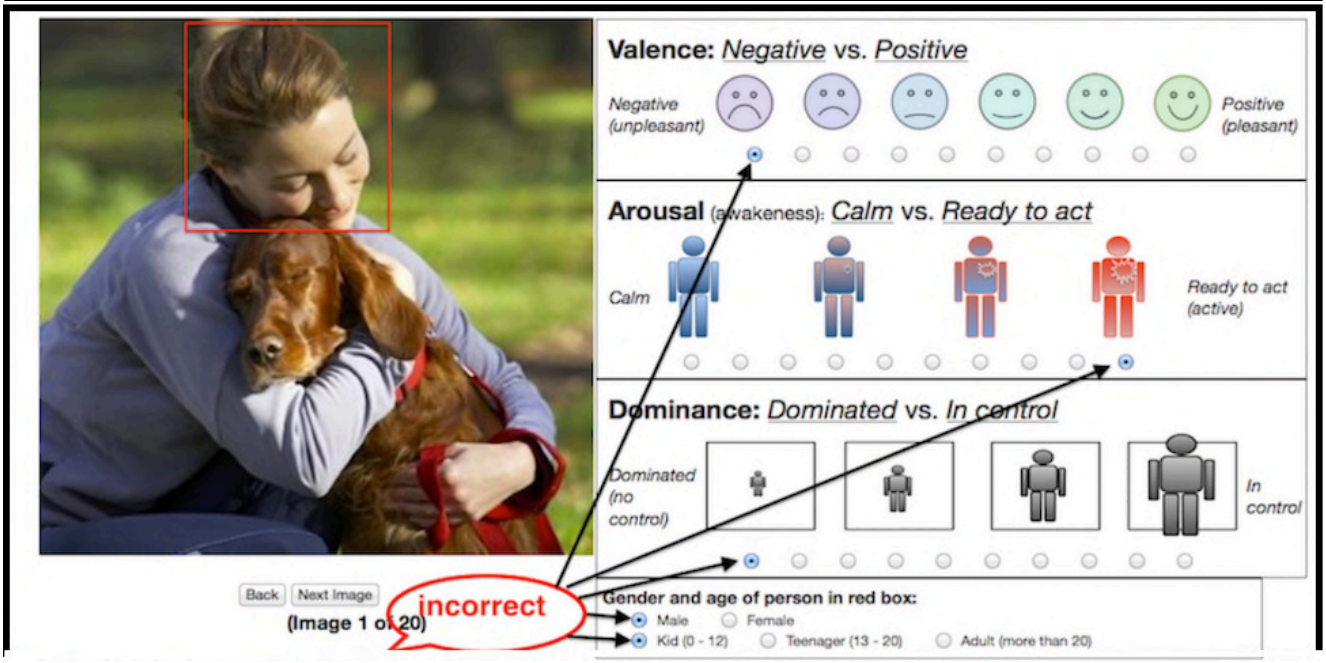


Figure 6: We define the 3 different dimensions for the workers in order to help the workers get the basic idea by showing examples. For each dimension, there is a representative image for a particular score. This helps to get acquainted of what the scores of each dimension mean.

Example:

- We have to label each of the three emotional dimensions for this person.
- First, we observe the picture: the person is hugging a dog, she looks happy and relaxed. The dog is quiet and seems to enjoy this hugging as well. The environment looks quiet, we can see some vegetation and we notice it is sunny, the weather looks nice.
- Now let us analyze each one of the emotional dimensions:
  1. **Valence** (*is the person having positive OR negative feelings?*): In this case (image below), she seems to be experiencing a positive emotion. So, her valence emotional level is positive. If we can imagine situations where a person could experience even higher positive emotion we should rate this dimension with a high score, but not with the highest possible.
  2. **Arousal** (*is the person calm (like sleepy) OR she feels energetic, agitated or excited and ready to act?*): In this case (image below), She seems relaxed and quiet, her body pose indicates that she is not too active, so her arousal scale is towards calm. We can imagine, however, situations where a person can feel even more calm (for instance someone almost sleeping in the sofa). For this reason we score this dimension towards calm, but not the most calm possible.
  3. **Dominance** (*does the person feel that the situation is not under her control OR she feels in control of the situation?*): In this case (image below), the situation is not threatening at all, so the person seems to be in control. So, in the dominance scale, her emotional state suggests a high level of control of the situation.
- Also select the appropriate age and the gender of the person in red box.



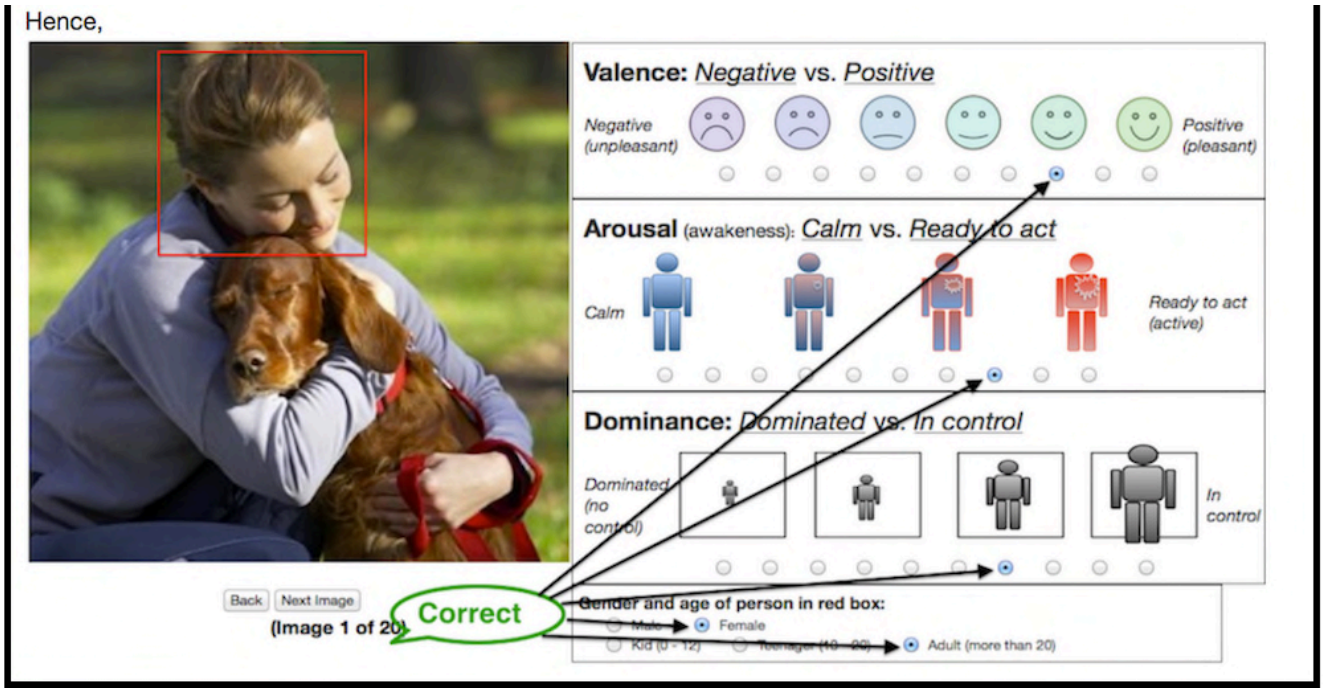


Figure 7: Showing the workers an example of how to annotate, meanwhile, explaining each step and giving reasons for selecting a particular score over other scores. We also give example of incorrect annotation to show that such instances are not good annotations. This is supposed to help them decide the score level while annotation. For this reason, we also include these instructions and examples at each person annotation.

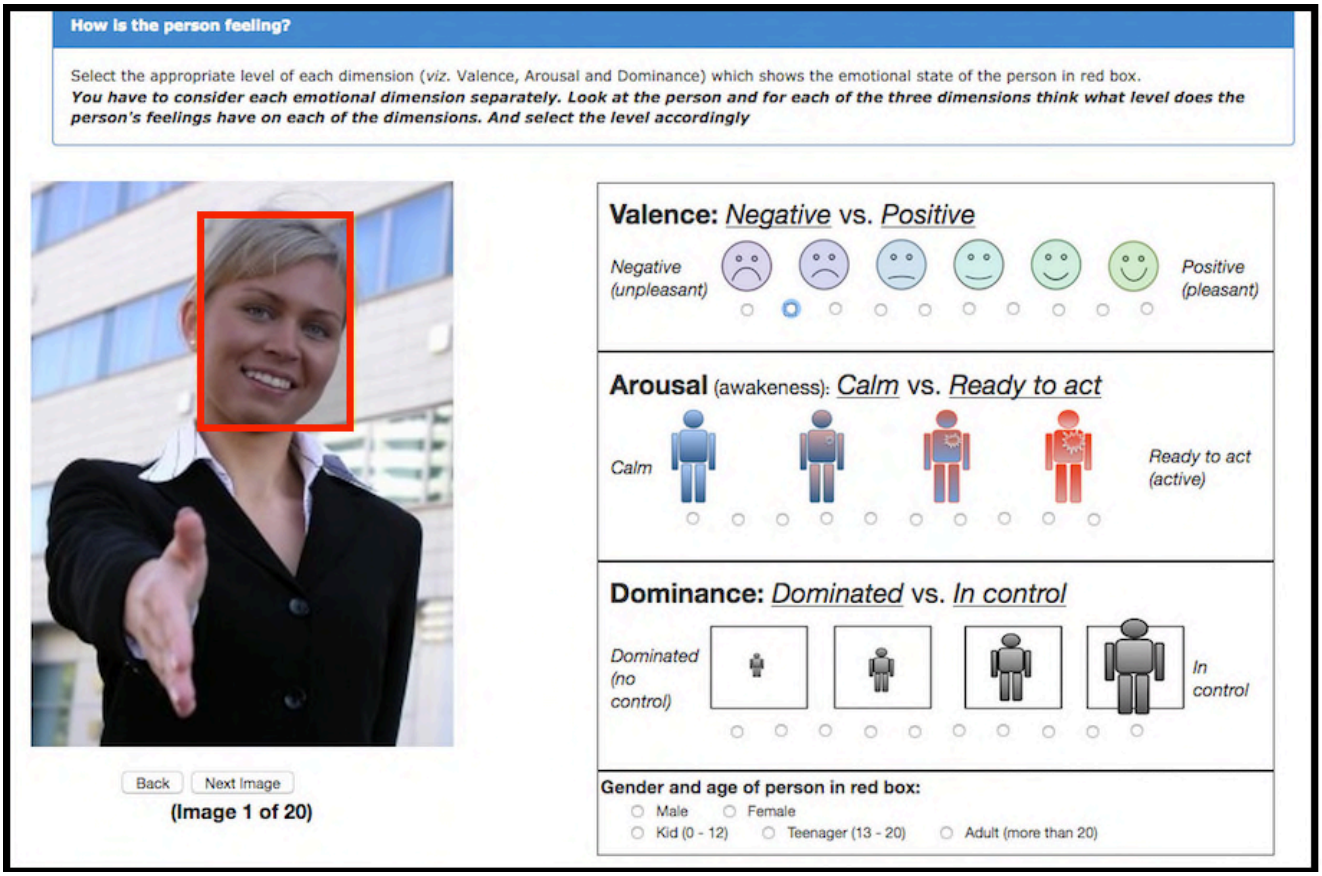


Figure 8: In the interface, the workers are shown all the 3 dimensions with the level of scores and asked to select appropriate level according to the person in the bounding box. We also ask the workers to estimate the gender and the age of the person in the bounding box

(3.1) Image Annotation

- Main Paper lines: $\mathbf{313 - 315}$

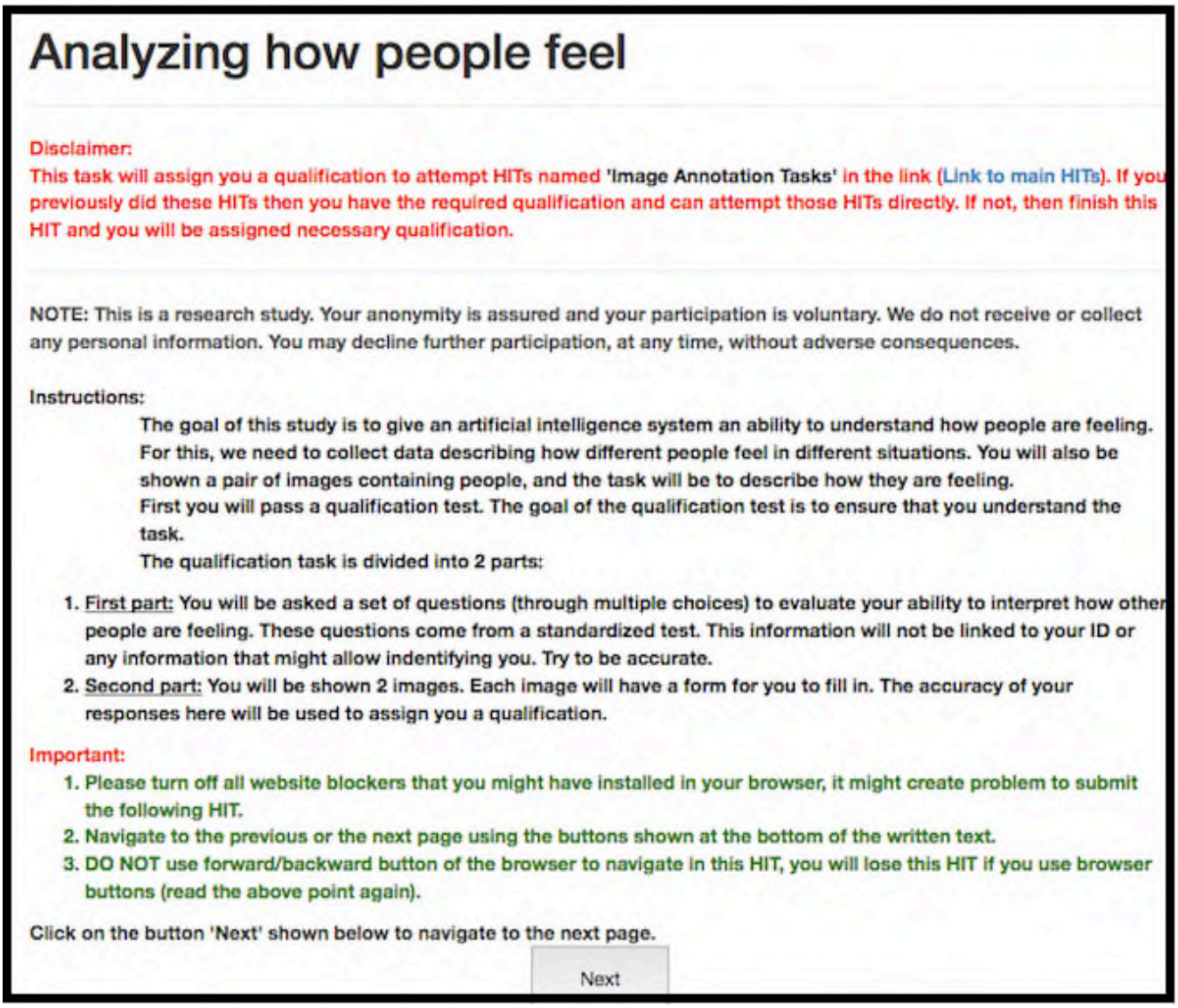


Figure 9: The first page of the EQ task containing an exhaustive and clear instructions and information. This helps the worker to get informed about the tasks ahead.

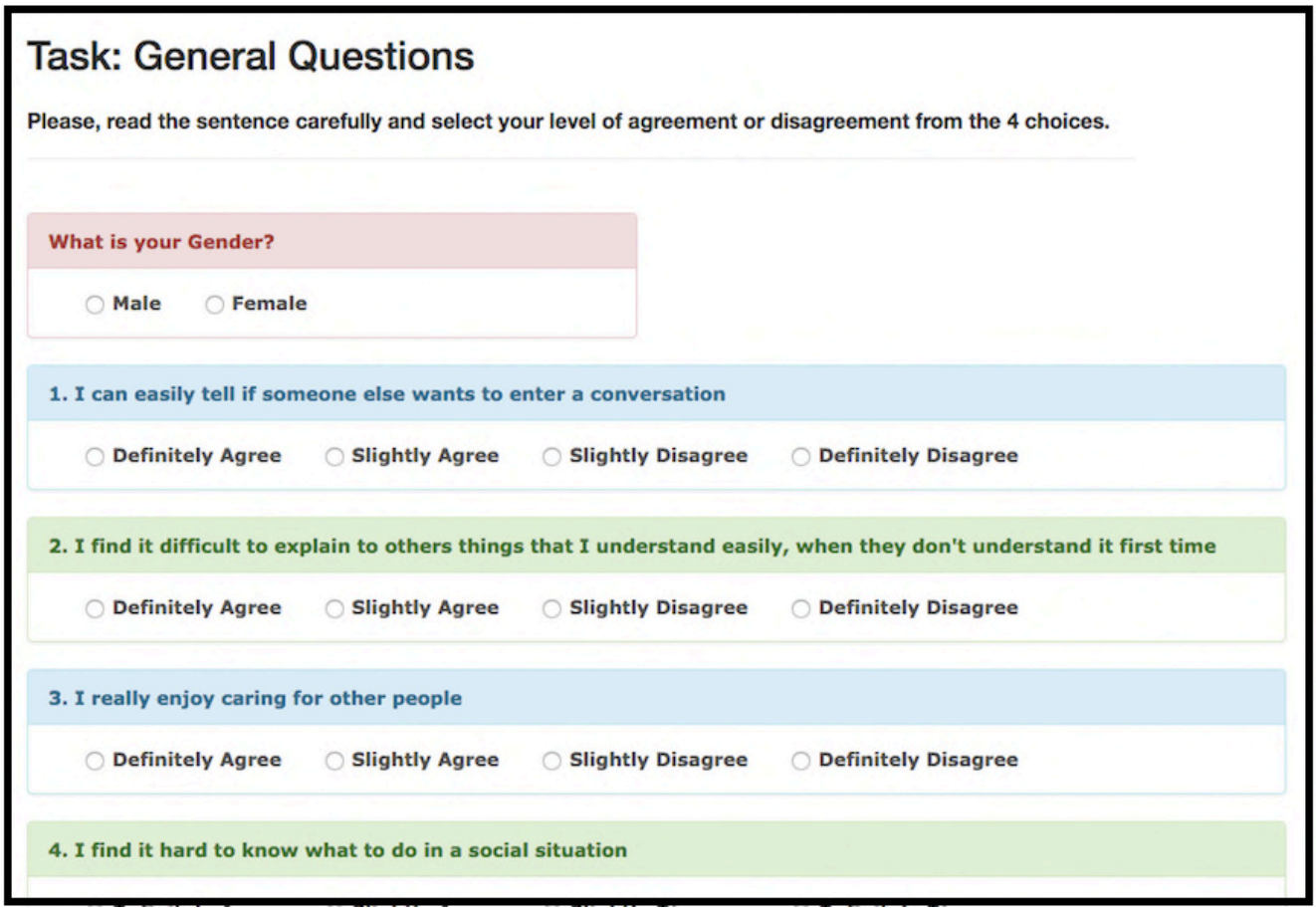


Figure 10: Snapshot of the standard questions asked to the workers to attempt [11]. These questions are very general in nature asking about certain situations a person might face in real life and the decisions, thereby, he/she takes. These decisions (or responses) help us estimate the empathizing quotient

Figure 11: We also inserted 2 control questions in the interface so that the worker does not attempt the questions randomly. These control questions are simple, clear and have a standard response. If the control questions are not correct then we present the worker with the warning message.

(3.2) Database Statistics

- Main Paper lines: $\mathbf{416 - 419}$

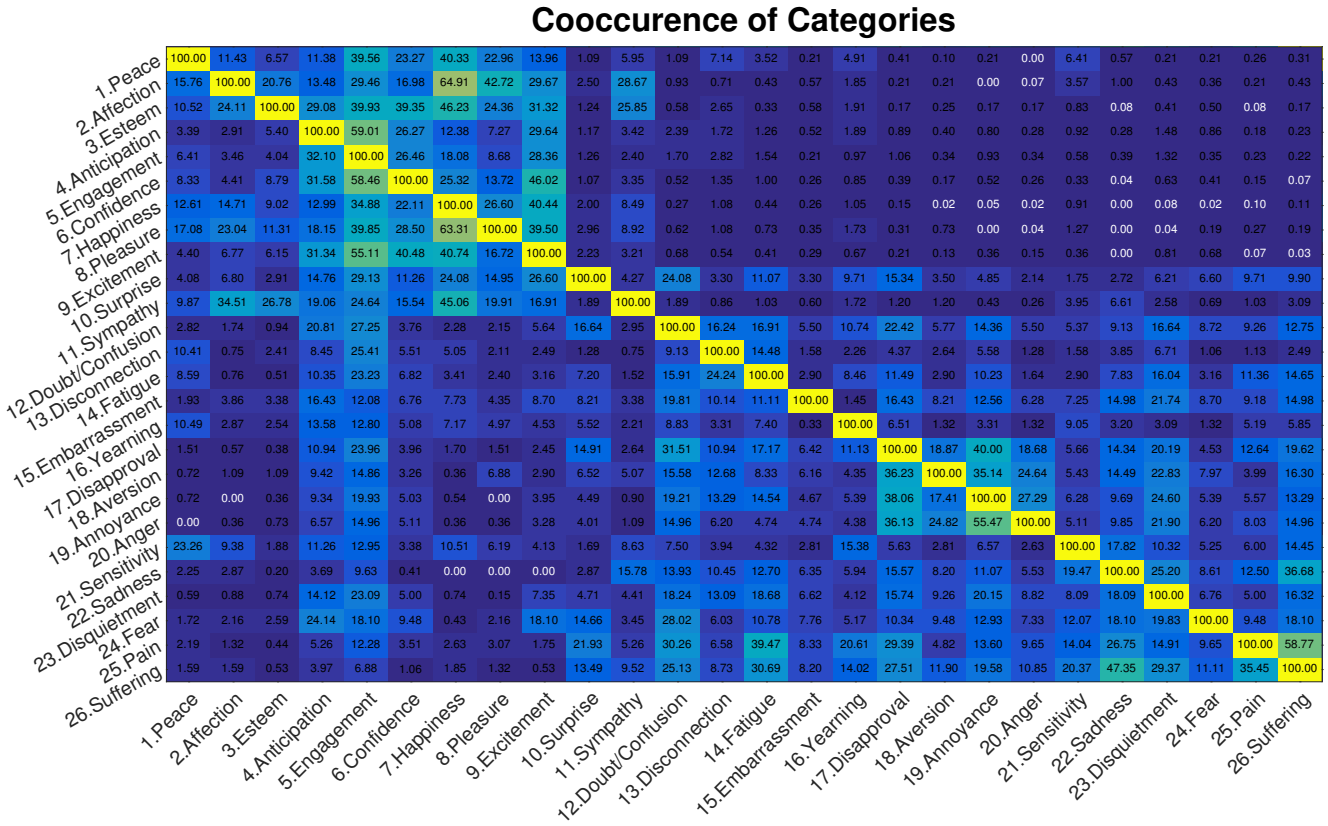


Figure 12: Co-occurence of Categories. Each row represents one feeling category. For each row, the column values represent the percentage of persons that are also annotated by the corresponding category of the column. ***Example.*** (7).Happiness has a huge overlap (65%) with row of (2).Affection, meaning that for all persons annotated with (2).Affection, 65% of them have been annotated with (7).Happiness as well.

(3.2) Database Statistics

- Main Paper lines: $\mathbf{476 - 478}$

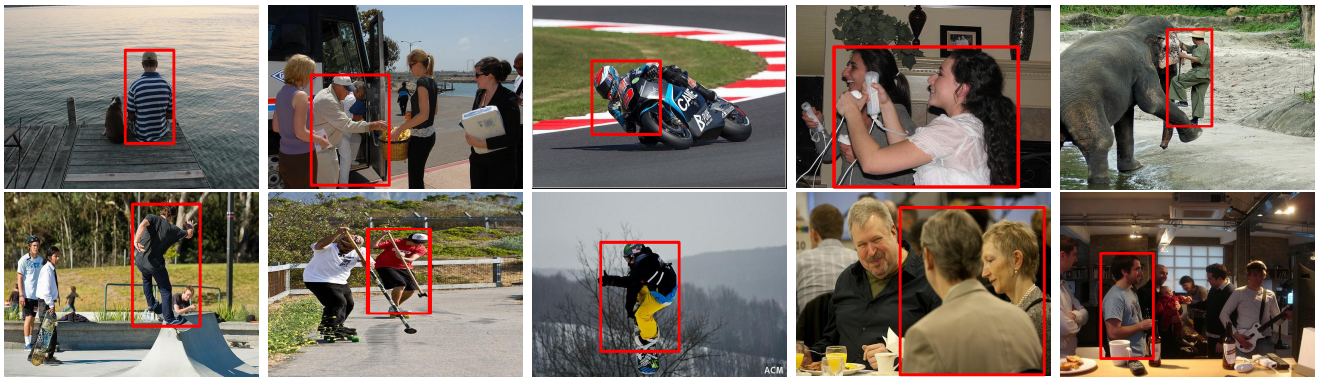


Figure 13: A sample set of images from our FEEL Database, which shows the variety of images where the person's face is obscured for various reasons. However, using contexts like pose, scene category, surrounding people's reactions and other object categories, it is easy to infer the feeling of the person.