Supplementary material: Emotional Features of Interactions with Empathic Agents

ANNEX: DETAILED DESCRIPTION OF THE STATISTICAL ANALYSES PERFORMED

Statistical analyses

Repeated measures ANOVA analyses were firstly performed on the data for each country separately and then on all countries data aggregated a single dataset. Each dimension was separately analysed. The following statistical analyses were considered:

1. Repeated measures ANOVAs to assess differences among annotators' labelling in terms of both categorical and dimensional labels. The within-subjects factors were the annotator and the time.

2. Repeated measures ANOVAs to assess changes in the assigned dimensional and categorical emotional dimensions across the five temporal segments, considering time as within-subjects factor and participants' gender as between-subjects factor.

3. Repeated measures ANOVAs to detect whether differences in the assigned labels' occurrence in each temporal segment was affected by the proposed scenario (generic and nutrition). The within-subject was the type of scenario and the participants' gender was considered as between-subjects factor.

Finally, with the aim to evaluate differences between Spain, Norway and France, repeated measures ANOVAs analyses were performed on the aggregated data, considering only the labelling provided by the first annotator in each country. For cross-countries comparisons, the following statistical analyses were considered:

1. Repeated measures ANOVAs to detect putative changes in the assigned emotional dimensional and categorical dimensions across the five temporal segments, considering the time as within-subjects factor and participants' gender and country as between-subjects factors.

2. Repeated measures ANOVAs to detect whether differences in the assigned labels' occurrences in each temporal segment was affected by the type of scenario (generic and nutrition) along which the user/EMPATHIC coach interaction was conducted. The scenario was considered as within-subject factor, whereas participants' gender and country were considered as between-factors.

For all the analyses the significant level was set at $\alpha = .05$ and differences among means were assessed though Bonferroni's post hoc tests.

In the following annex, only statistically significant results will be reported. It is worth mentioning that the repeated measures ANOVAs have been conducted only on the first annotator, in all the cases where no significant differences among annotators' labelling were observed; in the cases where significant differences among annotators' labelling were observed, the repeated measures ANOVAs have performed both on the first annotator's labelling and on the labelling provided by the second annotator (only on the occurrences of the labels which differ significantly). The results of post hoc comparisons report the averaged percentages.

Analyses of the Audio Data

Arousal in the audio

France

Significant differences emerged among annotators' labelling for "slightly excited" ($F(2,256)=6.421$; $p=.003$) and "neutral/calm"($F(2,256)=6.455$; $p=.003$) in the generic scenario. The second annotator (mean=32.88) perceived participants more frequently as "Slightly excited" compared to the first (mean=21.78) and third (mean=23.69) ones; whereas for the "Neutral/calm" label, the second annotator perceived participants less frequently as "calm"(mean=67.09) compared to the first (mean=78.22) and third (mean=76.31) ones. In the nutrition scenario, no differences among annotators' labellingwere observed for any of the labels' occurrence. According to the first and second annotators' labelling, no significant differences in any of the scenarios were found across temporal segments. Likewise, according to both annotators' labelling,

no significant differences between generic and nutrition scenarios at any temporal segments were observed.

Norway

The ANOVA results showed that there were no significant differences among the annotators' labelling in both scenarios. According to the first annotator's labelling no significant differences emerged among temporal segments for both scenarios and finally, the assigned labels' occurrence was not affected by the generic and nutrition scenarios.

Spain

The results showed no significant differences among annotators' labelling for both scenarios. The results revealed no significant changes along time for both scenarios. The comparison between the generic and nutrition scenarios revealed:

- the results showed an increase of the "slightly excited" label in the generic scenario with respect to the nutrition one in the second temporal segments $(F(1,65)=4.363, p=.041)$. Post hoc comparisons revealed that the participants were perceived more frequently as "slightly excited" in the generic scenario (mean=3.792) than in the nutrition one (mean=1.118).
- the results showed an increase of the "neutral/calm" label in the nutrition scenario with respect to the generic one in the second temporal segments $(F(1,65)=4.363, p=.041)$. Post hoc comparisons revealed that the participants were perceived more frequently as "neutral/calm" in the nutrition scenario (mean=98.882) than in the generic one (mean=96.208).

No other significant differences were observed.

Cross-countries comparisons

Results on the ANOVAs performed to detect modifications of the labels' occurrences along temporal segments revealed the following effects:

Generic scenario

- there was a significant difference among countries $(F(2,125)=38.331; p<<.01)$ on the occurrences of the label "slightly excited", according to which French participants (mean=21.50) were perceived more frequently as "slightly excited" compared to their Spanish (mean=2.60) and Norwegian (mean=3.92) counterparts.
- there was a significant difference among countries $(F(2,125)=27.016; p<<.01)$ on the occurrence of the label "neutral/calm". Post-hoc comparisons showed that French participants (mean=78.50) were perceived less frequently as "neutral" compared to Spanish (mean=96.57) and Norwegian (mean=96.09) ones.

Nutrition scenario

- a significant difference among countries $(F(2,125)=39.919; p<<.01)$ was observed on the "slightly excited" occurrences: French participants (mean=23,09) were perceived more frequently as "slightly excited" than Spanish (mean=1.47) and Norwegian (mean=2.94) ones.
- a significant difference among countries $(F(2,125)=30.327; p<<.01)$ was observed on the "neutral/calm" occurrences: French participants (mean=73.76) were perceived less frequently as "neutral" than Spanish (mean=96.51) and Norwegian (mean=96.99) ones.

Repeated measures ANOVAs performed to compare labels' occurrences between generic and nutrition scenarios showed no significant differences due to the type of scenarios.

Significant differences among countries were observed:

Label "Slightly excited"

During the first $(F(2,125)=41.740; p<<.01)$, the second $(F(2,125)=22.193; p=<<.01)$, the third $(F(2,125)= 32.602; p<<.01)$, the fourth $(F(2,125)=24.578; p<<.01)$ and the fifth $(F(2, 125)=51.849; p<<.01)$ temporal segments, post-hoc comparisons showed that French participants (means: T1=25.42; T2=18.29; T3=22.47; T4=19.31; T5=25.97) were always perceived more frequently as "slightly excited" compared to Spanish (means: T1=4.58; T2=2.46; T3=1.48; T4=1.36; T5=2.57) and Norwegian (means: T1=2.24; T2=3.40; T3=3.42; T4=2.54; T5=3.21) ones.

Label "Neutral/Calm"

In the first ($F_{(2,125)}$=35.114; p<<.01), the second ($F_{(2,125)}$=22.194; p=<<.01), the third ($F_{(2,125)}$= 23.182; p<<.01), the fourth ($F_{(2,125)}$=18.510; p<<.01) and the fifth ($F_{(2, 125)}$=30.617; p<<.01) temporal segments, post-hoc comparisons revealed that French participants (means: T1=71.95; T2=81.71; T3=81.71; T4=78.06; T5=72.73) were always perceived less frequently as calm compared to Spanish (means: T1=97.60; T2=97.55; T3=96.68; T4=97.46; T5=96.79) and Norwegian (means: T1=95,24; T2=96,60; T3=96.59; T4=96.28; T5=94.56) ones.

Dominance in the audio

France

Results showed that, in the generic scenario, the annotators' labelling did not significantly differ on the occurrence of the labels "rather dominant/controlling the situation" ($F_{(2,256)}$=2.287; p=.110) and "neither dominant nor intimidated" ($F_{(2,256)}$=2.563; p=.085). However, post-hoc comparisons showed that: the second annotator (mean=9.39) perceived participants significantly more frequently as in control of the situation compared to the third one (mean=5.81); the second annotator (mean=89.96) perceived significantly less frequently participants as neutral compared to the third one (mean=93.37). No significant differences among annotators' labelling were observed in the nutrition scenario. No significant differences across the temporal segments emerged from the analyses performed on the first annotator's labelling innone of the scenarios. Instead, according to the second annotator's labelling, the result showed significant changes alongtime:

Generic scenario

- Concerning the occurrence of the label "rather dominant/controlling the situation" ($F_{(4,124)}$=3.102; p=.018), participants were perceived more in control during the first temporal segment (mean=12.99) compared to the fourth one (mean=3.57)
- Concerning the occurrence of the label "neither dominant nor intimidated" ($F_{(4,124)}$=3.241; p=.014), participants were perceived less neutral during the first temporal segment (mean=86.15), compared to the fourth one (mean=96.43).

No significant differences were found in the labels' occurrence in the nutrition scenario. Concerning the repeated measures ANOVAs performed to test whether the assigned labels' occurrences differed due to the type of scenario, results on the first annotator's labelling showed that:

Label "rather dominant/controlling the situation"
- In the first temporal segments there was a difference due to the type of scenario ($F_{(1,31)}$=4.691; p=.038): in the generic scenario (mean=11.99) its occurrence was significantly higher than the nutrition one (mean=5.62). No other significant differences were observed.

According to the second annotator's labelling, results of ANOVAs showed the following significant results:

Label "rather dominant/controlling the situation"
- In the first temporal segments, there was a difference between scenarios ($F_{(1,31)}$= 4.732; p=.037): participants were perceived more frequently as in control in the generic scenario (mean=12.99) than the nutrition one (mean=5.50).
- In the fourth temporal segments, the occurrence of the label was affected by the type of scenario ($F_{(1,31)}$=5.848; p=.022): participants were perceived more frequently as in control during the nutrition scenario (mean=14,23) compared to generic one (mean=3.57).

Label "neither dominant nor intimidated"
- In the fourth temporal segments, the occurrence of the label was significantly different between scenarios ($F_{(1,31)}$=6.970; p=.013): it was higher in the generic scenario (mean=96.43) than nutrition one (mean=83.14).

Norway

The ANOVA results showed that there were no significant differences among the annotators' labelling in both scenarios. According to the first annotator's labelling no significant differences emerged among temporal segments for both scenarios and finally, the assigned labels' occurrence was not affected by the generic and nutrition scenarios.

Spain

No significant differences among annotators' labelling were observed in the generic scenario. Significant differences emerged among annotators' labelling for "rather dominant/controlling the situation" in the nutrition scenario (F(2,528)=4.442,p=.014). Post hoc comparisons revealed that the third annotator (mean=3.59) perceived participants less frequently as in control compared to the second annotator (mean=5.78). Significant effect emerged among annotators' labelling also for "rather intimidated/defensive" in the nutrition scenario (F(2,528)=5.408,p=.006). Post hoc comparisons revealed that the second annotator (mean=.40) perceived participants less frequently as defensive compared to the first annotator (mean=1.56).

Concerning the repeated measures ANOVAs performed to assess differences across the temporal segments, the results showed no significant changes for both scenarios.

Concerning the repeated measures ANOVAs performed to test whether the assigned labels' occurrences differed due to the type of scenario, according to the first annotator's labelling, the results showed significant difference between scenarios (F(1,65)=5.480,p=.022) in the occurrence of the label "rather intimidated/defensive": post hoc comparisons revealed that during the fourth temporal segments, the participants were perceived more frequently on the defensive in the nutrition scenario (mean=2.65) compared to the generic one (mean=.00). According to the second and third annotators' labelling, results showed no significant differences between scenarios.

Cross-countries comparisons

Results on the ANOVAs performed to detect modifications of the labels' occurrences along temporal segments revealed:

Generic scenario

- A significant difference among countries (F(2,125)=4.176; p=.018) on the occurrence of the "rather dominant/controlling the situation": French participants (mean=6.73) were perceived more frequently in control only with respect to Norwegian ones (mean=0.44).
- A significant interaction effect between time and country (F(8,500)=3.474; p=.001) on the occurrence of the "rather dominant/controlling the situation". The interaction effect reported that during the first temporal segment, French participants (mean=11.99) were perceived more frequently in control compared to their Spanish (mean=1.59) and Norwegian (mean=0.54) counterparts.
- A significant interaction effect between time and country (F(8,500)=3.382; p=.001) on the occurrence of the label "neither dominant nor intimidated". The interaction effect reported that during the first temporal segment, French participants (mean=88.01) were perceived less frequently as neutral compared to Spanish (mean=98.11) and Norwegian ones (mean=99.12).

No significant differences were found in the labels' occurrence in the nutrition scenario.

Repeated measures ANOVAs performed to compare labels' occurrences between generic and nutrition scenarios showed the following significant effects:

Label "Rather dominant/controlling the situation"

- During the first temporal segments, there was a significant difference among countries (F(2,125)=8.183; p<<.01) on the occurrence of the label: French participants (mean=8.81) were perceived more frequently as in control compared to Spanish (mean=2.45) and Norwegian (mean=1.25) ones.
- During the first temporal segments, there was an interaction effect between the type of scenario and the country (F(2,125)=4.078; p=.019) on the occurrence of the label. In the generic scenario, French participants (mean=11,99) were perceived more in control compared to their Spanish (mean=1.59) and Norwegian (mean=0.54) counterparts.
- During the fifth temporal segments, the occurrence of the label "rather dominant/controlling the situation" was affected by the country (F(2,125)= 3.244; p=.042): French participants (mean=8.30) were perceived more frequently in control with the respect to Norwegian (mean=0.98) ones.

Label "Neither dominant nor intimidated"

- During the first temporal segments, there was a difference among countries (F(2,125)= 9.302; p<<.01) on the occurrence of the label: French participants (mean=88.40) were perceived less frequently as neutral compared to their Spanish (mean=97.20) and Norwegian (mean=98.58).

- During the last temporal segments, there was a difference among countries ($F(2,125)= 3.124$; $p=.047$): post-hoc comparisons showed that French participants (mean=90.31) were perceived less frequently compared to their Norwegian (mean=98.91) counterparts.

Valence in the audio

France

Results of the ANOVAs to assess differences among annotator's labelling revealed a significant difference for the occurrence of the label "positive" in the generic scenario ($F(2,256)=3.804$; $p=.027$): the second annotator (mean=14.02) perceived participants less frequently as "positive" compared to the third one (mean=23.09).

In the nutrition scenario, there was a significant effect of the annotators' labelling for the occurrence of the label "neither positive nor negative" ($F(2,256)=4.014$; $p=.023$): the second annotator (mean=83.69) perceived participants more frequently as neutral compared to the third one (mean=77.74). Also in the nutrition scenario, the occurrence of "positive" label differed among annotators' labelling ($F(2,256)=4.675$; $p=.013$): the second annotator (mean=13.77) perceived less frequently participants as "positive" compared to the third one (mean=19.78).

Results of analyses performed to test differences along time showed the following significant differences:

Generic scenario, according to the first annotator's labelling:

- The occurrence of the label "neither positive nor negative" changed along temporal segments ($F(4,124)=3.988$; $p=.004$): during the first temporal segment (mean=65.16) participants were perceived less frequently as neutral compared to the second one (mean=83.06).
- The occurrence of the label "positive" changed along time ($F(4,124)=6.045$; $p<<.01$): in the second temporal segment (mean=13.70) it was lower compared to the first (mean=34.29) and to the fifth (mean=23.95) ones.

Generic scenario, according to the second annotator's labelling:

- The occurrence of the label "positive" in the generic scenario changed along time ($F(4,124)=9.800$; $p<<.01$): participants were perceived more frequently as "positive" during the first temporal segment (mean=26.18) with respect to the second (mean=9.69), third (mean=5.83) and fourth (mean=13.04) ones.

Nutrition scenario, according to the first annotator's labelling:

- The occurrence of the label "neither positive nor negative" changed across temporal segments ($F(2,124)=6.522$; $p<<.01$): in the last temporal segment (mean=66.04) participants were perceived less frequently as neutral compared to the second (mean=86.65), the third (mean= 84.63) and the fourth (mean=87.38) ones.
- The occurrence of the "positive" label changed due to time ($F(2,124)=8.670$; $p<<.01$): during the last temporal segment (mean=29.95) participants were perceived more frequently as "positive" compared to the second (mean=12.23), the third (mean=11.95) and the fourth (7.25) ones.

Nutrition scenario, according to the second annotator's labelling:

- The occurrence of the label "neither positive nor negative" significantly changed across temporal segments ($F(4,124)= 8.688$; $p<<.01$). Post-hoc comparisons revealed that during the last segment (mean=68.11) partici- pants were perceived less frequently as neutral compared to the second (mean=90.17), the third (mean=91.53) and the fourth (mean=87.26) ones.
- The occurrence of "positive" label changed due to time ($F(4,124)=8.166$; $p<<.01$): during the last segment (mean=28.56) participants were perceived more frequently as "positive" compared to the second (mean=9.17), the third (mean=8.24), and the fourth (mean=8.74) ones.

Repeated measures ANOVAs comparing the first annotator's labelling between the generic and nutrition scenarios showed the following significant effects:

Label "Positive":

- During the first temporal segments, the occurrence of the label significantly differed between scenarios ($F(1,31)=4.250$; $p=.048$): participants were perceived more frequently as "positive" in the generic scenario (mean=34.29) compared to the nutrition one (mean=22.99).

- During the fourth temporal segments, the occurrence of the label was significantly different between scenarios ($F(1,31)=7.675$; $p=.009$): it was higher in the generic scenario (mean=19.03) compared to the nutrition one (mean=7.25).

Norway

The ANOVAs results showed that there were no significant differences among the annotators' labelling in both scenarios. According to the first annotator's labelling no significant differences emerged among temporal segments for both scenarios and finally, the assigned labels' occurrence was not affected by the generic and nutrition scenarios. No significant differences emerged for participants' gender.

Spain

The ANOVAs results showed that there were no significant differences among the annotators' labelling in the generic scenario. Significant effect emerged among annotators' labelling for "positive" ($F(2,528)=6.041,p=.003$) and "neither positive nor negative" ($F(2,528)=6.013$, $p=.003$) in the nutrition scenario. Post hoc comparisons revealed that concerning "positive" label, the second annotator (mean=69.77) perceived participants less frequently as positive compared to the first annotator (mean=72.82), while concerning the "neither positive nor negative" label the second annotator(mean= 30.05) perceived participants more frequently as neutral compared to the first annotator (mean=25.04). Results of the analyses performed to detect modifications along temporal segments revealed:

Generic scenario, according to the first annotator's labelling:

- Concerning "positive" label ($F(4,260)=10.353,p<<.01$), post hoc comparisons revealed that this was due to the lower occurrences in the first temporal segment (mean=46.54) compared to the second (mean = 62.54) , third (mean=69.83), fourth (mean=68.78) and fifth (mean=64.11) ones.
- Concerning "neither positive nor negative" label ($F(4,260)=13.121,p<<.001$), post hoc test revealed that this was due to the higher occurrences in the first temporal segment (mean= 53.00) compared to the second (mean =37.27), third (mean=29.53), fourth (mean=29.24) and fifth (mean=34.65) ones.

Nutrition scenario, according to the first annotator's labelling:

- Concerning "positive" label ($F(4,260)=5.999,p<<.01$), post hoc comparisons revealed that this was due to the lower occurrences in the fourth (mean=69.67) and fifth temporal segments (mean=64.00) compared to the first (mean=76.416) and second (mean=79.285) ones.
- Concerning "neither positive nor negative" label ($F(4,260)=2.943,p=.021$), post hoc comparisons revealed that this was due to the higher occurrence in the fifth temporal segment (mean=31.13) compared to the second one (mean =20.56)

Nutrition scenario, according to the second annotator's labelling:

- Concerning "positive" label ($F(4,260)=15.649,p<<.01$), post hoc comparisons revealed that this was due to the lower occurrences in third (mean=68.08), fourth (mean= 64.02) and fifth temporal segments (mean=53.74) compared to the first (mean=75.34) and second (mean=79.26) ones.
- Concerning "neither positive nor negative" label ($F(4,260)=4.582,p=.001$), post hoc comparisons revealed that: occurrences were higher in the third temporal segment (mean= 30.635) compared to the second one (20.383) and they were higher in the fifth temporal segment (mean =41.902) compared to the first (mean= 24.049), second (mean=20.383) and third (mean=30.635) ones.

The repeated measures ANOVAs, performed to test whether the assigned labels' occurrences differed due to the type of scenario on the first annotator's labelling, showed:

Label "Positive":

- In the first temporal segments there was a difference between scenarios ($F(1,65)=58.764,p<<.01$). Post hoc comparisons revealed that the participants were perceived more frequently as "positive" in the nutrition scenario (mean=76.15) compared to the generic one (mean=46.54).

- In the second temporal segments, once again, there was a difference between scenarios $(F(1,65)=29.603, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as "positive" in the nutrition scenario (mean=79.29) compared to the generic one (mean=62.54).

Label "neither positive nor negative":

- In the first temporal segments there was a difference between scenarios $(F(1,65)=59.554, p<<.01)$. Post hoc comparisons revealed that the participants were perceived as more frequently as "neither positive nor negative" in the generic scenario (mean=53.00) compared to the nutrition one (mean=23.12).
- In the second temporal segments there was a difference between scenarios $(F(1,65)=29.305, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as "neither positive nor negative" in the generic scenario (mean=37.27) compared to the nutrition one (mean=20.56) .

According to the second annotator's labelling, the results showed significant differences between scenarios:

Label "Positive":

- In the first temporal segments there was a difference between scenarios $(F(1,65)=73.212, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as "positive" in the nutrition scenario (mean=75.34) compared to the generic one (mean=44.03).
- In the second temporal segments there was a difference between scenarios $(F(1,65)=52.837, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as "positive" in the nutrition scenario (mean=79.26) compared to the generic one (mean=58.21).

Label "neither positive nor negative":

- In the first temporal segments there was a difference between scenarios $(F(1,65)=73.833, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as neutral in the generic scenario (mean=55.50) compared to the nutrition one (mean=24.05).
- In the second temporal segments there was a difference between scenarios $(F(1,65)=51.935, p<<.01)$. Post hoc comparisons revealed that the participants were perceived more frequently as neutral in the generic scenario (mean=41.50) compared to the nutrition one (mean=20.38).

Cross-countries comparison

Results on the ANOVAs performed to detect modifications of the labels' occurrences along temporal segments revealed the following significant effects:

Generic scenario:

- There was a main difference among countries $(F(2,125)=117.054; p<<.01)$ on the occurrence of the label "positive", according to which Spanish participants (mean= 62.36) were associated with the highest occurrence of this label, followed by the French (mean= 22.13) and Norwegian (mean=2.93) ones .
- An interaction effect between time and country $(F(8,500)=8.778; p<<.01)$ was observed on the occurrence of the label "positive". During the first temporal segment, Spanish (mean=46.54) and French (mean=34.39) participants had a higher occurrence of this label with respect to Norwegian ones (mean=3.66). In all the other temporal segments, Spanish participants (means: T2=62.54; T3= 69.83; T4=68.78; T5=64.11) were perceived more frequentlyas "positive" compared to both Norwegian (means: T2=3.37; T3= 3.03; T4=0.32; T5= 4.02) and French (means: T2=13.70; T3= 19.68; T4=19.03; T5= 23.95) ones.
- There was difference among countries $(F(2,125)=117.371; p<<.01)$ on the occurrence of the label "neither positive nor negative", according to which Spanish participants (mean= 36.74) were associated with the lowest occurrence of this label, followed by the French (mean= 75.27) and Norwegian ones (mean=96.19).
- An interaction effect between time and country $(F(8,500)=8.564; p<<.01)$ was observed on the occurrenceof the label "neither positive nor negative". During the first temporal segments, Spanish (mean=53.00) and French (mean=65.16) participants had a lower occurrence of this label with respect to Norwegian participants (mean=96.07). In the second temporal segment, Norwegian (mean=95.59) and French participants (mean=83.06) were perceived more frequently as neutral with respect to Spanish participants (mean=37.27). In all the other temporal segments, Norwegian participants (means: T3= 96.70; T4=97.98; T5=94.60) were perceived more frequently

as neutral compared to both French (means: T3= 76.04; T4=80.22; T5=71.87) and Spanish (means: T3= 29,53; T4=29,24; T5=34,65) ones.

- There was a significant difference among countries (F(2,125)=4.444; p=.014) on the occurrence of the label "negative": French participants (mean=2.60) were perceived more frequently as "negative" with respect to their Spanish(mean=0.21) counterparts.

Nutrition scenario:

- The occurrence of the label "positive" significantly differed among countries (F(2,125)=187.022; p<<.01). Post-hoc comparisons revealed that Spanish participants (mean=72.13) had the highest occurrence of this label, followed by French (mean=16.87) and Norwegian (mean=3.04) ones.
- An interaction effect between time and country (F(8,500)=7.135; p<<.01) on the occurrence of the label "positive" was observed. During the first and fifth temporal segments, Norwegian participants (means: T1=3.28; T5=5.48) were perceived less frequently as "positive" compared to both Spanish (means: T1=76.41; T5=64.00) and French (means: T1=22.99; T5=29.95) ones. In the other temporal segments, Norwegian (means: T2=1.98; T3=1.83; T4=2.64) and French (means: T2=12.23; T3=11.95; T4=7.25) were perceived less frequently as "positive" with respect to their Spanish counterparts (means: T2=79.29; T3=71.26; T4=69.67).
- There was a difference among countries (F(8,500)=188.183; p<<.01) on the occurrence of the label "neither positive nor negative": Norwegian participants (mean=96.11) were perceived more frequently as neutral compared totheir French (mean=79.17) and Spanish (mean=25.15) counterparts.
- An interaction effect between time and country (F(8,500)=5.965; p<<.01) on the occurrence of the label "neither positive nor negative". During the first and fifth temporal segments, Norwegian participants (means: T1=95.66; T5=93.81) were perceived more frequently as neutral compared to both Spanish (means: T1=23.12; T5=31.12) and French (means: T1=71.18; T5=66.04) ones. In the other temporal segments, Norwegian (means: T2=97.90; T3=97.44; T4=95.76) and French (means: T2=86.65; T3=84.63; T4=87.38) participant were perceived more frequently as neutral with respect to their Spanish counterparts (means: T2=20.56; T3=24.74; T4=26.22).

Repeated measures ANOVAs performed to compare the assigned labels' occurrences between generic and nutrition scenarios showed the following significant effects:

Label "Positive":

- A significant difference among countries was observed in the first (F(2,125)=83.278; p<<.01), the second (F(2,125)=179.506; p<<.01), the third (F(2,125)=133.206; p<<.01); the fourth (F(2,125)=157.226; p<<.01) andthe last (F(2,125)= 187.022; p<<.01) temporal segments: Spanish participants were always associated with the highest occurrence of the label (means: T1=61.48; T2=70.91; T3=70.55; T4=69.22; T5=64.06), followed by French (means: T1=28.64; T2= 12.96; T3=15.82; T4=13.14; T5=29.95) and Norwegian (means: T1=3.47; T2=2.68; T3=2.57; T4=1.48; T5=4.75) ones.
- In the first temporal segments, a significant interaction effect between the type of scenario and country (F(2,125)= 28.607; p<<.01) was observed. In the generic scenario, Spanish (mean=46.54) and French (mean=34.29) participants had significant higher occurrence of the label with respect to their Norwegian counterparts (mean=3.66);whereas in the nutrition scenario Spanish participants' label occurrence (mean=76.42) was higher compared to bothNorwegian (mean=3.28) and French (mean=22.99) ones.
- In the second temporal segments, there was a significant interaction effect between the type of scenario and country (F(2,125)=9.810; p<<.01): in both scenarios, Spanish participants (means: generic=62.54; nutrition=79.29) were associated with the highest occurrence of the label, followed by French (means: generic= 13.70; nutrition= 12.23) and Norwegian (means: generic=3.37; nutrition=1.98) ones.

Label "Neither positive nor negative":

- A significant difference among countries was found in the first (F(2,125)=79.653 p<<.01), the second (F(2,125)=173.984; p<<.01), the third (F(2,125)=131.128; p<<.01); the fourth (F(2,125)=158.387; p<<.01) andthe last (F(2,125)=102.427; p<<.01) temporal segments: Spanish participants were always associated with the lowest occurrence of the label (means: T1=38.06; T2=28.92; T3=27.13; T4=27.73; T5=32.89), followed by French(means: T1=68.19; T2= 84.86; T3=80.34; T4=83.80; T5=68.95) and Norwegian (means: T1=95.86; T2=96.74; T3=97.07; T4=96.87; T5=94.21) ones.
- During the first temporal segments, the interaction effect between the type of scenario and the country (F(2,125)=20.620; p<<.01) was significant. In the generic scenario, Spanish (mean=53.00) and French (m=65.16)

participants were perceived less frequently as neutral with respect to their Norwegian counterparts (mean=96.07); whereas in the nutrition scenario, Spanish participants (mean=23.12) significantly differed from both Norwegian (mean=95.66) and French (mean=71.18) ones in the label's occurrences.

- During the second temporal segments, the interaction effect between the type of scenario and the country $(F(2,125)=10.790; p<<.01)$ was significant: both in the generic and nutrition scenarios, Spanish participants (means: generic=37.27; nutrition=20.56) were perceived less frequently as neutral compare to both French (means: generic=83.06; nutrition=86.65) and Norwegian ones (means: generic=95.59; nutrition=97.90), which did not significantly differ from each other.

Label "Negative":

- During the third temporal segments, there was a significant interaction effect between the type of scenario and country $(F(2,125)=4.197; p=.017)$. The significant difference was due to the generic scenario, in which French participants (mean=4.28) were perceived more frequently as "negative" with respect to their Spanish counterparts (mean=0.23).

Categorical Emotion perceived in the audio

France

Results of the analyses performed on the occurrence of the emotion labels in the generic scenario revealed no significant differences among annotators' labelling $(F(2,256)=2.788; p=.069)$ in the occurrence of the "calm" label. However, post-hoc comparisons revealed that the first annotator (mean=94.37) perceived participants significantly more frequently as "Calm" compared to the third one (mean=92.50).

The effect of the annotators' labelling $(F(2,256)=2.463; p=.093)$ on the occurrence of label "amused" in the generic scenario was not significant. However post-hoc comparisons showed that the first annotator (mean=2.93) perceived participants less frequently as "amused" compared to the third one (mean= 3.25).

In the nutrition scenario, no differences among annotators' labelling were observed for any of the labels' occurrence. Results of analyses testing the differences in labels' occurrence along the temporal segments showed that, according to both the first and the third annotators' labelling, there were no significant effects due to time in none of the scenarios.

The analyses comparing the occurrence of the labels between the generic and nutrition scenarios revealed that, according to the first and the third annotators' labelling, no differences were found due to the type of scenario.

Norway

The ANOVAs results showed that there were no significant differences among the annotators' labelling in the generic and nutrition scenarios.

The results showed a significant change among temporal segments in the generic scenario:

- Concerning the occurrences of "calm" label $(F(4,116)=4.556,p=.002)$. Post hoc comparisons revealed that this was due to the higher occurrences in the second temporal segment (mean=97.78) compared to the fifth one (mean=89.83).
- Concerning the occurrences of "amused" label $(F(4,116)=5.615,p<<.01)$. Post hoc comparisons revealed that this was due to the lower occurrences in the second temporal segment (mean=1.03) compared to the fifth one (mean=10.15).

The results showed a significant change among temporal segments in the nutrition scenario:

- Concerning the occurrences of "calm" label $(F(4,116)=4.225,p=.003)$. Post hoc comparisons revealed that this was due to the higher occurrences in the second temporal segment (mean=98.85) compared to the fifth one (mean=91.68).

According to the first annotator's labelling, repeated measures ANOVAs comparing the labels' occurrence between the generic and nutrition scenarios showed no significant differences due to the type of scenario for any of the labels.

Spain

Significant effect emerged among annotators' labelling for "puzzled" labels ($F_{(2,520)}=4.409$; $p=.014$). Post-hoc comparisons showed that the third annotator (mean=2.65) perceived participants more frequently as "puzzled" compared to the second one (mean=2.24). Also in the nutrition scenario, significant effect emerged among annotators' labelling for "puzzled" labels ($F_{(2,528)}=6.268$), $p=.003$). Post-hoc comparisons showed that the third annotator (mean=3.37) perceived participants more frequently as "puzzled" compared to the second one (mean=2.68). The results showed a significant change among temporal segments:

Generic scenario, according to the first annotator's labelling:

- Concerning the occurrences of "amused" label ($F_{(4,260)}=2.866$, $p=.024$), post-hoc comparisons revealed that in the fifth temporal segment the participants were perceived more frequently as "amused" (mean=4.47) compared to the fourth one (mean=1.13).
- Concerning the occurrences of "puzzled" label ($F_{(4,260)}=3.393$, $p=.010$), post-hoc comparisons revealed that in the fifth temporal segment the participants were perceived less frequently as "puzzled" (mean=1.05) compared to the second (mean=3.97) and fourth (mean=3.67) ones.

Generic scenario, according to the third annotator's labelling:

- Concerning the occurrences of "puzzled" label ($F_{(4,256)}=3.179$, $p=.015$), post-hoc comparisons revealed that in the fifth temporal segment the participants were perceived less frequently as "puzzled" (mean=.94) compared to the second (mean=4.33) and fourth ones (mean=3.40).

Nutrition scenario, according to the first annotator's labelling:

- Concerning the occurrences of "calm" label ($F_{(4,260)}=6.279$, $p<<.01$), post-hoc comparisons revealed that in the first (mean=96.57) and second temporal segments (mean=95.76) the participants were perceived more frequently as "calm" compared to the third (mean=90.16), fourth (mean=89.78) and fifth (mean=87.47) ones.
- Concerning the occurrences of "amused" label ($F_{(4,260)}=4.296$, $p=.002$), post-hoc comparisons revealed that in the fifth temporal segment the participants were perceived more frequently as "amused" (mean=4.813) compared to the first (mean=1.543) and third ones (mean=1.069).

According to the second annotator's labelling, the results showed no significant change across time for the "puzzled" labels.

The repeated measures ANOVAs, performed to test whether the assigned labels' occurrences differed due to the type of scenario on the first annotator's labelling, showed a significant differences between scenarios:

Label "Amused":

- In the fourth temporal segments ($F_{(1,65)}=4.591$, $p=.036$) the participants were perceived more frequently as "amused" in the nutrition scenario (mean=2.95) compared to the generic one (mean=1.13)

Label "Puzzled":

- In the second temporal segments ($F_{(1,65)}=5.670$, $p=.020$), the participants were perceived more frequently as "puzzled" in the generic scenario (mean=3.97) compared to the nutrition one(mean=1.86).
- In the third temporal segments ($F_{(1,65)}=7.410$, $p=.008$), the participants were perceived more frequently as "puzzled" in the nutrition scenario (mean=5.13) compared to the generic one (mean=1.73).

According to the second annotator's labelling, the results showed a significant differences between scenarios for "puzzled" labels:

- In the third temporal segments ($F_{(1,65)}=5.738$, $p=.019$), the participants were perceived more frequently as "puzzled" in the nutrition scenario (mean=5.19) compared to the generic one (mean=2.22)

Cross-countries comparison

Results on the ANOVAs performed to detect modifications of the labels' occurrences along temporal segments showed the following significant differences:

Generic scenario:

- There was an interaction effect between time and country ($F_{(8,500)}=3.501$; $p=.001$) on the occurrence of the label "amused". This was due to the last temporal segment, during which Norwegian participants (mean=10.25) were perceived more frequently as "amused" compared to both French (mean=2.72) and Spanish (mean=4.47) ones.
- There was a significant difference due to the country ($F_{(2,125)}=4.419$; $p=.014$) on the occurrence of the label "puzzled". Post-hoc comparisons showed that Norwegian participants (mean=0.58) were perceived less frequently as "puzzled" compared to their French (mean=2.68) and Spanish (mean=2.52) counterparts.

Nutrition scenario:

- There was a significant difference among countries ($F_{(2,125)}=7.838$; $p=.001$) on the occurrence of the label "puzzled". Post-hoc comparisons showed that Norwegian participants (mean=0.33) were perceived less frequently as "puzzled" compared to their French (mean=2.11) and Spanish (mean=3.00) counterparts.

Repeated measures ANOVAs performed to compare the assigned labels' occurrences between generic and nutrition scenarios among countries showed the following significant effects:

Label "Calm":

- During the second temporal segments, there was a difference among countries ($F_{(2,125)}=3.320$; $p=.039$. Post-hoc comparisons revealed that Norwegian participants (mean=98.31) were perceived more frequently as "calm" with respect to their Spanish counterparts (mean=94.56).

Label "Amused":

- During the third temporal segments, there was a difference among countries ($F_{(2,125)}=4.529$; $p=.013$). Post-hoc comparisons showed that Norwegian participants (mean=3.86) were perceived more frequently as "amused" compared to both Spanish (mean=1.70) and French (mean=1.16) ones.
- During the fifth temporal segments, there was a difference among countries ($F_{(2,125)}=4.006$; $p=.021$). Post-hoc comparisons showed that Norwegian participants (mean=9.16) were associated with higher occurrence of the label compared to Spanish (mean=4.64) and French (mean=4.16) ones.

Label "Puzzled":

- A significant differences among countries on the occurrence of the label was observed in the third ($F_{(2,125)}=4.196$; $p=.017$), in the fourth ($F_{(2,125)}=6.593$; $p=.002$) and in the fifth ($F_{(2,125)}=4.364$; $p=.015$) temporal segments. Post-hoc comparisons revealed that during the third and fourth temporal segments, Norwegian (means: T3=0,29; T4=0,30) participants were perceived less frequently as "puzzled" with respect to their Spanish (means: T3=3,43; T4= 3,41) counterparts. During the last temporal segments, Norwegian (mean=0.08) participants significantly differed from both Spanish (mean=1.94) and French (mean=2.56) participants.

Annotation of Video Data through Categorical Emotional Labels

France

Results of the repeated measures ANOVAs testing differences among annotators' labelling showed no significant effects in the generic scenario. In the nutrition scenario, the occurrence of the label "neutral" differed due to the annotators' labelling ($F_{(1,128)}=5.102$; $p=.031$). Post-hoc comparisons revealed that the first annotator (mean=84.71) perceived participants lessfrequently as neutral compared to the second one (mean=88.56).

According to the first annotator's labelling, repeated ANOVAs performed to detect differences along temporal segments revealed that in the generic scenario:

- the occurrence of the label "amused" changed along time ($F_{(4,124)}=4.239$; $p=.003$): it was significantly higher during the first temporal segment (mean=11.27) with respect to the second one (mean=4.47).
- the occurrence of the label "pensive" significantly differed along temporal segments ($F_{(4,124)}=5.038$; $p=.001$): participants were perceived less frequently as "pensive" during the first temporal segment (mean=8.60) compared to the second (mean=20.92) and the fourth (mean=18.97) ones.

According to both annotators' labelling, no differences along temporal segments were observed in the nutrition scenario.

Concerning the repeated measures ANOVAs performed to detect differences in the labels' occurrences between generic and nutrition scenario, the results of the first annotators' labelling revealed:

Label "Surprised":

- A significant difference due to the type of scenario (F(1,31)=4.607; p=.040) during the fourth temporal segments: in the generic scenario (mean=0.16) participants were perceived more frequently as "surprised" compared to the nutrition one (mean=0.00).

Label "Amused":

- Significant differences due to the type of scenario during the first (F(1,31)=9.594; p=.004); the fourth (F(1,31)=8.323; p=.007) and the fifth (F(1,31)=5.706; p=.023) temporal segments. Occurrences were always significantly higher in the generic scenario (means: T1=11.27; T4=5.83; T5=10.01) compared to the nutrition one (means: T1=4.77; T4=2.03; T5=5.00).

Label "Pensive":

- Significant differences due to the type of scenario during the second (F(1,31)=4.412; p=.044), the third (F(1,31)=7.792; p=.009), the fourth (F(1,31)=6.208; p=.018) and the fifth (F(1,31)=13.364; p=.001) temporal segments. In details, the label's occurrence was higher in the generic scenario (means: T2=20.92; T3=19.41; T4=18.97; T5=16.59) with respect to the nutrition one (means: T2=14.00; T3=9.00; T4=9.42; T5=3.99).

Label "Neutral", according to the first annotator's labelling:

- Significant differences due to the type of scenario during the third (F(1,31)=6.759; p=.014); the fourth (F(1,31)=10.038; p=.003) and the fifth (F(1,31)=16.472; p<<.01). Participants were perceived less frequently as "neutral" in the generic scenario (means: T3=72.62; T4=73.79; T5=72.66) compared to the nutrition one (means: T3=84.89; T4= 86.52; T5=88.36).

Label "Neutral", according to the second annotator's labelling:

- The occurrences of this label were significantly different between generic and nutrition scenarios in the first (F(1,31)=6.904; p=.013), the second (F(1,31)=9.943; p=.004), the third (F(1,31)=13.936; p=.001), the fourth (F(1,31)=17.747; p<<.01) and the fifth (F(1,31)=22.589; p<<.01) temporal segments. In details, participants were perceived less frequently as "neutral" during the generic scenario (means: T1=80.85; T2= 74.92; T3=75.15; T4=73.58; T5=73.57) with respect to the nutrition one (means: T1=87.77; T2= 85.24; T3=88.49; T4=90.61; T5=91.82).

Norway

Significant effects emerged among annotators' labelling in the generic scenario:

- Concerning the "pensive" labels (F(1,120)=27.547,p<<.01), post-hoc comparisons showed that the second annotator (mean=16.63) perceived the participants more frequently as "pensive" compared to the first one (mean=14.00).
- Concerning the "neutral" labels (F(1,120)=10.296,p=.003), post-hoc comparisons showed that the first annotator (mean=76.34) perceived the participants more frequently as "neutral" compared to the second one (mean=73.94).

Significant effect emerged among annotators' labelling in the nutrition scenario:

- Concerning the "pensive" labels (F(1,120)=6.963,p=.013), post-hoc comparisons showed that the second annotator(mean=21.02) perceived the participants more frequently as "pensive" compared to the first one (mean=19.41).

Generic scenario, according to the first annotator's labelling:

- Concerning the occurrences of "amused" label (F(4,116)=4.883,p=.001), post hoc comparisons revealed that in the fifth temporal segment the participants were perceived more frequently as "amused" (mean=10.51) compared to the fourth one (mean=4.08).
- Concerning the occurrences of "pensive" label (F(4,116)=6.921,p<<.01), post hoc comparisons revealed that in the first temporal segment (mean=7.12) participants were perceived less frequently as "pensive" compared to the second (mean=16.77), third (mean=19.17) and fourth (mean=18.62) ones. Furthermore, post-hoc comparisons revealed that in the fifth temporal segment (mean=11.99) the participants were perceived less frequently as "pensive" compared to the fourth one (mean=18.625).

- Concerning the occurrences of "neutral" label (F(4,116)=3.888,p=.005), post-hoc comparisons revealed that in the first temporal segment (mean=83.35) the participants were perceived more frequently as "neutral" compared to the third one (mean=72.09).

Generic scenario, according to the second annotator's labelling:

- Concerning the occurrences of "pensive" label (F(4,116)=8.188,p<<.01), post-hoc comparisons revealed that in the first temporal segment (mean = 7.75) the participants were perceived less frequently as "pensive" compared to the second (mean=19.456), third (mean=22.405) and fourth ones (mean=22.131).
- Concerning the occurrences of "neutral" label (F(4,116)=4.828,p=.001), post-hoc comparisons revealed that in the first temporal segment the participants were perceived more frequently as "neutral" (mean=82.736) compared to the second (mean=73.165) and third ones (mean=69.853).

Nutrition scenario, according to the first annotator's labelling:

- Concerning the occurrences of "amused" label (F(4,116)=6.326,p<<.01), post-hoc comparisons revealed that in the fifth temporal segment (mean=11.35) the participants were perceived more frequently as "amused" compared to the second (mean=3.33) and third (mean=4.50) ones.
- Concerning the occurrences of "pensive" labels (F(4,116)=7.217,p<<.01), post-hoc comparisons revealed that in the fifth temporal segment (mean=11.34) the participants were perceived less frequently as "pensive" compared to the first (mean=18.276), second (mean=24.166), third (mean=22.150) and fourth (mean=20.771) ones.

Nutrition scenario, according to the second annotator's labelling:

- Concerning the occurrences of "pensive" label (F(4,116)=6.294,p<<.01), post-hoc comparisons revealed that in the fifth temporal segment (mean=14.63) the participants were perceived less frequently as "pensive" compared to the second (mean=26.05), third (mean=23.76) and fourth (mean=22.46) ones.

The repeated measures ANOVAs, performed to test whether the assigned labels' occurrences differed due to the type of scenario on the first annotator's labelling, showed:

Label "pensive":

- In the first temporal segments (F(1,29)=16.189,p<<.01), post hoc comparisons revealed that the participants were perceived more frequently as "pensive" in the nutrition scenario (mean=18.28) compared to the generic one (mean=7.12).
- In the second temporal segments (F(1,29)=5.194,p=.030), post hoc comparisons revealed that the participants were perceived more frequently as "pensive" in the nutrition scenario (mean=24.17) compared to the generic one (mean=16.77).

Label "neutral":

- In the first temporal segments (F(1,29)=5.429,p=.027), post hoc comparisons revealed that the participants were perceived more frequently as "neutral" in the generic scenario (mean=83.35) compared to the nutrition one (mean=74.46).

According to the second annotator's labelling, the results showed a significant differences between scenarios:

Label "pensive":

- In the first temporal segment (F(1,29)=16.223,p<<.01), post hoc comparisons revealed that the participants were perceived more frequently as "pensive" in the nutrition scenario (mean=18.84) compared to the generic one (mean=7.75).

Label "neutral":

- In the first temporal segments (F(1,29)=5.884,p=.022), post hoc comparisons revealed that the participants were perceived more frequently as "neutral" in the generic scenario (mean=82.74) compared to the nutrition one (mean=73.95).

Spain

Significant effect emerged among annotators' labelling concerning the "surprised" labels:

- In the generic scenario $(F(1,264)=5.745, p=.019)$, post hoc comparisons revealed that the second annotator (mean=.15) perceived the participants more frequently as "surprised" compared to the first one (mean=.04).
- In the nutrition scenario $(F(1,264)=4.749, p=.033)$, post hoc comparisons revealed that the second annotator (mean=0.23)perceived the participants more frequently as "surprised" compared to the first one (mean=0.02).

The results showed a significant change across time:

Generic scenario

- Concerning "pensive" labels $(F(4,260)=17.758, p<<.01)$: the results showed the participants were perceived more frequently as "pensive" in the second (mean=18.02), third (mean=21.17) and fourth temporal segments (mean=17.37) compared to the first one (mean=8.990); and the participants were perceived less frequently as "pensive" in the fifth temporal segment (mean=11.728) compared to the second (mean=18.02), third (mean=21.17) andfourth (mean=17.37) ones.
- Concerning the "neutral" labels $(F(4,260)=14.944, p<<.01)$: the results showed the participants were perceived less frequently as "neutral" in the second (mean=80.64), third (mean=77.68) and fourth temporal segments (mean=81.55) compared to the first one (mean=89.00); and the participants were perceived more frequently as "neutral" in the fifth temporal segment (mean=86.31) compared to the second (mean=80.64), third (mean=77.68) and fourth (mean=81.55) ones.

Nutrition scenario

- Concerning the "pensive" labels $(F(4,260)=8.227, p<<.01)$: the results showed the participants were perceived less frequently as "pensive" in the fifth temporal segment (mean=9.19) compared to the first (mean=17.344), second (mean=20.29) and third (mean=22.63) ones. In the fourth temporal segment (mean 12.611) occurrences were lower compared to the second (mean=20.29) and third (mean=22.63) ones.
- Concerning the "neutral" labels $(F(4,260)=8.784, p<<.01)$: the results showed the participants were perceived more frequently as "neutral" in the fifth temporal segment (mean=88.955) compared to the first (mean=81.331), second (mean=79.15) and third (mean=81.31) ones. In the fourth temporal segment (mean=85.76) occurrences were higher compared to the second one (mean=79.15).

The repeated measures ANOVAs, performed to test whether the assigned labels' occurrences differed due to the type of scenario on the first annotator's labelling, showed:

Label "pensive":

- In the first temporal segments $(F(1,65)=18.777, p<<.01)$, post hoc comparisons showed that the participants were perceived more frequently as "pensive" in the nutrition scenario (mean=17.34) compared to the generic one (mean=8.99).
- In the fourth temporal segments $(F(1,65)=6.188, p=.015)$, post hoc comparisons showed that the participants were perceived more frequently as "pensive" in the generic scenario (mean=17.37) compared to the nutrition one (mean=12.61).

Label "neutral":

- In the first temporal segments $(F(1,65)=14.707, p<<.01)$, post hoc comparisons showed that the participants were perceived more frequently as "neutral" in the generic scenario (mean=89.00) compared to the nutrition one (mean=81.33).
- In the fourth temporal segments $(F(1,65)=4.500, p=.038)$ the participants were perceived more frequently as "neutral" in the nutrition scenario (mean=85.76) compared to the generic one (mean=81.55).

Cross-countries comparisons

Results on the repeated measures ANOVAs performed to detect modifications of the labels' occurrences along temporal segments among countries revealed that:

Generic scenario:

- There was a difference among countries (F(2,125)=11.313; p<<.01) in the occurrence of the label "amused". Post-hoc comparisons revealed that Spanish participants (mean=1.47) were perceived less frequently as "amused" compared to both French (mean=7.69) and Norwegian (mean=6.60) ones.
- The interaction effect between time and country (F(8,500)=3.647; p=.001) on the occurrence of the label "amused" was significant. In details, during the first temporal segments, Spanish participants (mean=1.94) were associated with lower occurrence of the label with respect to French ones (mean= 11.27); whereas in the third, fourth and fifth segments, Spanish participants (means: T3=1.08; T4=1.07; T5=1.95) were perceived significantly less frequently as "amused" compared to both their French (means: T3=6.90; T4=5.83; T5= 10.01) and Norwegian (means: T3= 6.46; T4= 4.08; T5= 10.51) counterparts.
- There was a significant difference among countries (F(2, 125)=3.438; p=.035) in the occurrence of the label "neutral". Post-hoc comparisons revealed that Spanish participants (mean= 83.04) were perceived more frequently as "neutral" with respect to their French counterparts (mean= 74.40).

Nutrition scenario:

- There was a difference due to country (F(2,125)=15.223; p<<.01) in the occurrence of the label "amused". Post-hoc comparisons showed that this was due to Spanish participants (mean=1.17) which were associated with lower occurrences of the label compared to their French (mean=3.46) and Norwegian (mean=6.46) counterparts.
- The occurrence of the label "amused" was affected by the interaction effect between time and country (F(8,500)=3.958; p<<.01). In details, during the first and second temporal segments, Spanish participants (means: T1=1.32; T2=0.56) were perceived less frequently as "amused" compared to both their French (means: T1=4.77; T2=2.68) and Norwegian (means: T1=7.23; T2=3.33) counterparts. During the last three temporal segments, both Spanish (means: T3= 0.61; T4=1.58; T5=1.80) and French (means: T3=2.83; T4=2.03; T5=5.00) participants were perceived less frequently as "amused" with respect to Norwegian ones (T3=4.51; T4=5.89; T5=11.35).
- The occurrence of the label "pensive" was significantly different among countries (F(2,125)=4.742; p=.010): French participants (mean=8.87) were perceived less frequently as "pensive" compared to both Norwegian (mean=19.34) and Spanish (mean=16.41).
- The occurrence of the label "neutral" significantly differed among countries (F(2,125)=4.440; p=.014). Post-hoc comparisons showed that Norwegian participants (mean=74.11) were perceived less frequently as "neutral" compared to their French (mean= 84.90) and Spanish (mean=83.30) counterparts.

Repeated measures ANOVAs performed to compare the assigned labels' occurrences between generic and nutrition scenarios among countries showed the following significant effects:

Label "Amused":

- Significant differences due to country were observed in the first (F(2,125)= 10.302; p<<.01), the second (F(2,125)=6.687; p=.002), the third (F(2,125)=10.652; p<<.01), the fourth (F(2,125)=6.785; p=.002), and the fifth (F(2,125)=12.716; p<<.01) temporal segments. Post-hoc comparisons revealed that Spanish participants (means: T1=1.63; T2=0.94; T3=0.85; T4=1.32; T5=1.88) were always perceived less "amused" compared to both French (means: T1=8.02; T2=3.58; T3=4.86; T4=3.93; T5=7.51) and Norwegian (means: T1=7.17; T2= 4.10; T3=5.48; T4=4.99; T5=10.93) ones.
- Significant interaction effects between the type of scenario and the country emerged in the first (F(2,125)=9.280; p<<.01), the fourth (F(2,125)=8.975; p<<.01), and the fifth (F(2,125)=4.515; p=.013) temporal segments. In details, during the first temporal segments, in the generic scenario Spanish participants (mean=1.94) were perceived less frequently as "amused" with respect to French ones (mean=11.27); whereas in the nutrition scenario, Spanish participants (mean= 1.32) were associated with lower occurrences of the label compared to both French (mean=4.77) and Norwegian (mean= 7.23) ones. During the fourth temporal segments, in the generic scenario Spanish participants (mean=1.07) were associated with a lower occurrence of the label compared to both French (mean=5.83) and Norwegian (mean=4.08) ones; whereas in the nutrition scenario the difference was due to Norwegian participants (mean=5.89), which were perceived more frequently as "amused" with respect to both French (mean=2.03) and Spanish (mean=1.58) ones. During the fifth temporal segments: in the generic scenario, French (mean=10.01) and Norwegian (mean=10.51) participants were perceived more frequently as "amused" compared to their Spanish counterparts (mean=1.95); whereas, in the nutrition scenario, Norwegian participants (mean=11.35) were perceived more frequently as "amused" compared to both French (mean= 5.00) and Spanish (mean=1.80) ones.

Label "Pensive":

- Significant interaction effects between the type of scenario and country were observed in the first ($F_{(2,125)}=5.845$; p=.004), the third ($F_{(2,125)}=3.543$; p=.032), the fourth ($F_{(2,125)}=3.519$; p=.033) and the fifth ($F_{(2,125)}=7.229$; p=.001) temporal segments. In details, during the first temporal segments, the significant difference was due only to the nutrition scenario: French participants (mean=7.95) were perceived less frequently as "pensive" compared to both Spanish (mean=17.34) and Norwegian (mean=18.28) ones. Also during the third temporal segments, the significant difference concerned only the nutrition scenario: in this case, French participants (mean=9.00) were perceived less frequently with respect to Spanish ones (mean=22.63). During the fourth and fifth temporal segments, the difference was significant in the nutrition scenario, in which French participants (means: T4=9.42; T5=3.99) were associated with a lower occurrence of the label with respect to the Norwegian ones (means: T4=20.77; T5=11.34).

Label "Neutral":

- Significant interaction effects between type of scenario and country were found in the first ($F_{(2,125)}=8.448$; p<<.01), the fourth ($F_{(2,125)}=4.429$; p=.014) and the fifth ($F_{(2,125)}=6.885$;p=.001) temporal segments. More specifically, during the first temporal segments: in the generic scenario French participants (mean=79.13) were perceived less "neutral" with respect to their Spanish (mean=89.00) counterparts; whereas in the nutrition scenario, French participants (mean=85.40) were perceived less "neutral" compared to their Norwegian counterparts (mean=74.46). During the fourth temporal segments, the difference was significant only in the nutrition scenario: Norwegian participants (mean=73.34) were perceived less frequently compared to both French (mean=86.52) and Spanish (mean=85.76) ones. Finally, during the fifth temporal segments, in the generic scenario, French (mean=72.66) and Norwegian (mean=75.25) participants were perceived less frequently as neutral compared to Spanish ones (mean=86.31). In the nutrition scenario, Norwegian participants (mean=76.98) were associated with lower occurrence of the label compared to French (mean=88.36) and Spanish (mean=88.96) ones.