1 **A 'spoon full of sugar' helps the medicine go down: how a participant friendly**

2 **version of a psychophysics task significantly improves task engagement,**

3 **performance and data quality in a typical adult sample.**

4

5 Marie L. Smith[1], M. Letizia Cesana[1,2], Emily K. Farran[3], Annette Karmiloff-Smith[1] &

6 Louise Ewing[1,4,5]

7

8 [1]Psychological Sciences, Birkbeck College, University of London, Malet Street, London,

9 WC1E 7HX

10 [2] *Department of Psychology, Catholic University of the Sacred Heart, Largo Gemelli 1,*

11 *20123 Milan, Italy.*

12 [3]Psychology and Human Development, UCL Institute of Education, University College

13 London, 25 Woburn Square, London WC1H 0AA

14 [4]School of Psychology, University of East Anglia, Norwich Research Park, NR4 7TK

15 [5] ARC Centre of Excellence in Cognition and its Disorders, School of Psychology,

16 University of Western Australia, 35 Stirling Hwy, Perth, 6009

17

18 **Abstract**

19   Few would argue that the unique insights brought by studying the typical and

20 atypical development of psychological processes are essential to building a

21 comprehensive understanding of the brain. Often, however, the associated challenges of

22 working with non-standard adult populations results in the more complex

23 psychophysical paradigms being rejected as too complex. Recently we created a child

24 (and clinical group) friendly implementation of one such technique – the reverse

25 correlation Bubbles approach and noted an associated performance boost in adult

26 participants. Here, we compare the administration of three different versions of this

27 participant-friendly task in the same adult participants to empirically confirm that

28 introducing elements in the experiment with the sole purpose of improving the

29 participant experience, not only boost the participant's engagement and motivation for

30 the task but results in significantly improved objective task performance and stronger

31 statistical results.

32

33

Introduction

There can be little doubt that necessity is the mother of invention, and the driving force for considering a methodological approach in a new light. Reverse correlation methods have a long productive history across a diverse range of topics in psychological and biological sciences (Ahumada & Lovell, 1971; Marmarelis & Marmarelis, 1978). Relatively recently they have been applied to the specific topic of face perception (e.g. Haig, 1995, Gosselin & Schyns, 2001, Sekuler, Gaspar, Gold & Bennett, 2004) and provided some important insights to the understanding of this vital ability. However, while approaches such as these have gleaned a wealth of information from healthy adult participants (typically the classic undergraduate student sample (Heinrich, Heine & Norenzayan, 2010)), their technical requirements have largely precluded a more general applicability that encompasses children and most atypical groups. To address this, we developed a participant-friendly version of one such technique (*Bubbles,* Gosselin & Schyns, 2001) and for the first time were able to use this approach successfully to better understand the development of face processing in typical children (6-12yrs: Ewing et al, 2017a, Ewing et al, 2017b) and adults with a neurodevelopmental disorder (Williams syndrome, see Ewing et al, 2017c).

In a standard adult *Bubbles* experiment, participants are expected to complete a large number of trials to guarantee comprehensive sampling of the stimulus space. Typically, this would be many hundreds of trials (at least 500 per condition, often more, e,g. Gosselin & Schyns, 2001, Smith et al, 2005), completed over multiple, extensive testing sessions. In adapting the paradigm to a non-standard audience we faced two important challenges: firstly, ensuring that we had sufficient information sampling to perform the Bubbles analysis and secondly ensuring that our participants remained fully engaged and motivated for as long as possible. To address the former point, rather than test a small number of individuals over many trials (as is typical), we tested large numbers of individuals over a relatively small number of trials. To address the latter point we introduced a number of modifications to the testing sessions, including shorter blocks with an onscreen countdown block progress bar, an interactive and encouraging experimenter sitting alongside the participant and engaging with them during all

2

67   breaks, and finally the introduction of the puzzle-bubble game during breaks. This game

68   involved the participants guessing the name of famous films/locations/tv-shows from

69   as little visual information as possible; cheeky bubbles 'hid' the key details, but could be

70   removed by the experimenter to provide further clues. Anecdotally, these changes and

71   the puzzle bubble game in particular, appeared surprisingly popular for children and

72   adults alike!

73   Although the effects of mental fatigue are well-known to negatively impact

74   cognitive performance (e.g. Boksem, Meijman & Lorist, 2005; Hopstaken, van der

75   Linder, Bakker & Koppier, 2015) with underlying changes in brain activation patterns

76   (e.g. Lorist, Boksen & Ridderinkof, 2005; Boksem, Meijman & Lorist, 2006; Borghini et

77   al, 2012; Tanaka, Ishii & Watanabe, 2014) there tends to be only minimal consideration

78   of the participant experience during the administration of repetitive tasks often asked

79   of participants in Psychology experiments. Mental fatigue occurs as a result of

80   sustained periods of demanding task performance and is typically characterized by

81   changes in mood and motivation (e.g. Boksem & Tops, 2008), and in particular a

82   reduction in task engagement (Hopstaken, van der Linden, Bakker & Kompier, 2015a).

83   Due to its importance in driving workplace errors and accidents, the study of mental

84   fatigue has often focused on the practical implications for occupational settings.

85   However, as mental fatigue is often directly linked to brain processes critical for

86   performance in psychophysical tasks (e.g. attention, Boksem, Meijman, Lorist, 2005,

87   global/local processing bias, van der Linder & Eling, 2006, executive control, van der

88   Linder, Frese, & Meijman, 2003) it follows that by overlooking their impact, researchers

89   of human behaviour may be deleteriously adding noise to their studies.

90   Research suggests that one way to counter the effects of mental fatigue is to

91   boost the rewards associated with participation (Boksem & Tops, 2008; Hopstaken, van

92   der Linder, Bakker & Kompier, 2015a; Hopstaken, van der Linder, Bakker & Kompier,

93   2015b) to re-engage fatigued participants in a given task. Given this, we were interested

94   to observe if our participant-friendly task modifications, which were specifically

95   designed to engage young / cognitively impaired individuals in our demanding,

96   repetitive and relatively boring tasks, could also have a measurable impact on task

97   performance and data quality in a standard adult sample.

98   We set out to validate the impact and effectiveness of our task engagement

99   strategy and the modifications made to the operation of the task by running three

100   identical base versions of the paradigm with the same adult participants in a single

101   testing session. In one version adults performed the task with no experimenter

102   interaction during the entirety of the task. There was no puzzle bubble game and only

103   generic self-paced "take a break" screens between blocks. In a second version, adults

104   again performed the experiment with no experimenter interaction, but with the puzzle

105   bubble game (played independently) separating blocks (even numbered blocks only).

106   Finally, in the third version, the experimenter interacted with the participant as they

107   played the puzzle bubble game (matching the participant-friendly implementation). All

108   other aspects of the methodology remained constant across the three versions of the

109   task. Furthermore, we employed a modified short-form of the Intrinsic Motivation

110   Inventory (IMI, Ryan, 1982) to directly assess each participant's subjective experience

111   of each experimental condition to determine if our manipulations significantly altered

112   the participants' experience of completing the task.

113   We directly compared performance across the different versions of the task, with

114   the expectation that the introduction of both the puzzle-bubble game (to enforce breaks

115   between trial blocks thus ensuring that the blocks are spaced out and to alleviate the

116   tedium of completing many similar trials) and interaction with the observer during

117   breaks would lead to better performance on the task and cleaner statistical results.

118   Comparing version 2 and version 1 permits us to evaluate the effectiveness of

119   the puzzle bubble game in boosting task engagement in itself, while the comparison of

120   version 2 and version 3 establishes the extent to which any improved performance is

121   driven by interaction with the experimenter. Direct comparison between self-report

122   measures of task engagement (from the IMI) and objective performance metrics (from

123   the Bubbles task) allows us to explicitly establish if greater task engagement is

124   significantly tied to experimental outcomes on a psychophysics task such as this.

125   To the best of our knowledge, this is the first time that the impact of the

126   participant experience has been explored in the context of a repetitive visual

127   psychophysics task conducted under typical experimental testing conditions (not those

128   designed to specifically induce mental fatigue by having participants perform the same

129   task repeatedly for a number of hours with no breaks). Should the subjective participant

130   experience and task engagement directly impact cognitive performance and resulting

131   data quality, then there are clear implications across a wide range of research areas in

132   the Psychological Sciences.

133

Methods

Participants

      30 adults (10male, mean age = 26.2, SD = 10.1) completed a single testing session lasting approximately 45minutes. All participants had normal or corrected to normal vision, no history of psychological problems, and provided signed informed consent. The study was approved by the ethics board of the Department of Psychological Sciences, Birkbeck College, University of London.

Procedure

      Using a repeated measures design, in a single testing session participants completed three bubbles conditions. Each took 10 to 15 minutes to complete, and were identical except for the introduction of the puzzle-bubble game during breaks between blocks (versions 2 and 3), and standardised interaction with the experimenter during the puzzle bubble game (version 3). Participants each completed a single puzzle bubble challenge per break (for a total of 4 challenges across the 512 trials of task versions 2 and 3), with each challenge lasting approximately 3 mins. The order in which participants completed each version of the task was randomised via a Latin square procedure with ten participants completing each order of the different versions.

      In the Bubbles task, participants were asked to categorize sub-sampled versions of expressive faces by the expression shown. The approach works by presenting only some parts of a stimulus (typically visual) to the participant on each trial and relating categorization decisions to the information that was presented. On each trial, most of the stimulus is hidden from view and only the information located behind a number of randomly positioned circularly symmetric Gaussian apertures is made available to the participant to make their categorization decisions. The location of the apertures varies randomly across trials so that over sufficient trials an exhaustive search of the visual space will have been conducted. Reverse correlating the location of the apertures with categorization responses permits the experimenter to establish which visual regions are significantly correlated with categorization performance and therefore can be concluded to be essential for the task at hand.

Stimuli were fearful, happy, angry and sad expressions taken from the California Facial Expression database (CAFÉ, as used in previous Bubbles expression categorization

166   studies, e.g. Smith et al, 2005; Smith & Merluscal, 2014; Schyns, Petro & Smith, 2007). As

167   per existing Bubbles studies of facial expression categorization, stimuli were

168   decomposed in 6 non-overlapping spatial frequency bands (SF) of one octave each

169   (120-60, 60-30, 30-15, 17-7.5, 7.5-3.8 cycles per image). To create a single experimental

170   stimulus, each SF band was sampled independently with randomly positioned Gaussian

171   apertures (the Bubbles) whose size was adjusted at each scale to reveal 3 cycles per

172   aperture and whose number (per scale) was adjusted to ensure equivalent sampling of

173   each SF scale (i.e. more small high SF bubbles than the larger low SF bubbles). The

174   sampled information from each scale was then recombined into a single stimulus image

175   comprising visual information across the SF bands (see Gosselin & Schyns, 2001 and

176   Smith et al, 2005 for fuller details of the stimulus generation process). The total number

177   of apertures (Bubbles) over all SF scales was adjusted on a trial per trial basis via a

178   staircase algorithm to target a performance criterion of 75% correct. To this end, poor

179   performance resulted in more information on a subsequent trial (i.e. more bubbles),

180   while higher than target performance resulted in a reduction in the amount of

181   information presented on subsequent trials (i.e. less bubbles).

182   In each version of the task participants completed 512 emotion categorization

183   trials (128 per emotion) by categorizing each stimulus by emotion (labelled keyboard

184   keys denoted fearful, happy, angry, sad and don't know), for a total of 1536 trials over

185   the course of the full experiment comprising the three task versions. A short practice

186   phase prior to testing confirmed that participants could correctly categorize the non-

187   Bubbled (i.e., intact) face stimuli by expression and introduced the participants to the

188   response keys. Participants sat 70cm from the experiment which ran on MATLAB using

189   the Psychophysics toolbox (Pelli), such that stimuli subtended a viewing angle of 5.36 x

190   3.7degrees of visual angle.

191   Unlike standard implementations of the Bubbles technique, in the modified

192   child-friendly version, we added a 'don't know' response to prevent participants from

193   correctly guessing and adding unnecessary noise to the data (don't know responses

194   were coded as incorrect). Furthermore, we introduced a count-down bar onscreen that

195   permitted participants to gauge their position in a block, and reduced the length of

196   individual blocks to a few minutes (64 trials) rather than around 5minutes.

197   To gauge interest/motivation, participants completed a short form of the

198   intrinsic motivation inventory assessment (IMI, Ryan, 1982) at the end of each

199 experimental condition. In this questionnaire we asked participants to rate (on a scale
200 of 1 to 7) how they felt about the task they had just completed in terms of their interest
201 and enjoyment (2 separate questions), their perceived competence, the effort they put
202 into their performance, the importance to them of doing well, the degree of pressure
203 they felt, how related they felt to the experimenter and finally how important they felt
204 the task was.
205
206 **Results**
207 **Bubbles results**
208 We considered two performance metrics as dependent measures: the amount of
209 information (i.e. number of bubbles) required to achieve the target performance of 75%
210 correct for each emotion and the actual percentage correct achieved (NB. with a small
211 number of trials it is not possible to perfectly stabilize performance at the target 75%
212 correct), see Figure 1A. Alongside this, we examined the quality of the bubbles solution
213 i.e. the visual information that is significantly associated with categorization of each
214 emotional expression. A one way repeated measures ANOVA with task version (1,2,3)
215 as the within subjects factor indicated a clear main effect on the amount of information
216 required to achieve good performance levels ($F(2,58) = 3.8$, $p=0.029$, $\eta^2=0.12$). Planned
217 comparisons revealed that participants required significantly less information for task
218 version 3 ($M = 85$ bubbles) compared to task version 2 ($M = 97$bubbles, $F(1,29) = 5.6$,
219 $p=0.025$, $\eta^2=0.16$), but there was no such drop in number of bubbles for task version 2
220 compared to task version 1 ($M = 93$ bubbles, $F(1,29) = 0.9$, $p=0.35$, $\eta^2=0.03$). An
221 equivalent ANOVA on percentage correct scores indicated a trend for a main effect of
222 condition here too ($F(1.3, 37.6) = 3.4$, $p=0.06$, $\eta^2=0.11$), with planned comparisons again
223 showing that participants are performing slightly better in task version 3 (74.4%)
224 compared to task version 2 (72%, $F(1,29) = 4.2$, $p=0.049$, $\eta^2=0.13$), but with no
225 improvement for task version 2 compared to task version 1 (73.4%, $F(1,29) = 2.3$,
226 $p=0.14$, $\eta^2=0.07$).
227 To evaluate the effectiveness of the task version manipulations on the quality of
228 the bubbles solution we considered the information processing results for the two most

229   well researched emotional expression categorizations: fear and happiness[1]. The critical

230   visual information for both fear and happiness categorizations has been confirmed

231   across a number of studies in typical adult participants. For fearful categorizations the

232   crucial visual information has been repeatedly shown to comprise the wide-open eyes

233   across scales in higher spatial frequencies (scales 1-3), alongside the open mouth

234   (scales 3 and 4, e.g. Smith & Merlusca, 2015; Smith et al, 2005, F. Smith & Schyns, 2009;

235   Adolphs et al, 2005). For happiness categorisations it is the wide-open mouth, from fine

236   detail in the higher spatial frequencies through to the broad low spatial frequency

237   mouth shape information (Smith & Merlusca, 2015; Smith et al, 2005, F. Smith & Schyns,

238   2009; Adolphs et al, 2005).

239   For both fear and happy, and all three task version scenarios, the bubbles

240   solution replicates *most[2]* of the key features of these established processing profiles.

241   Figure 1B shows only those regions that pass the corrected statistical tests ($p<0.05$,

242   Chauvin et al, 2005) highlighted on a sample face. Significant regions observed under

243   task version 1 are coded in red, those from task version 2 in green and finally those of

244   task version 3 in blue. Note that where the same regions were significant in multiple

245   task versions it is colour coded in the RGB colour space combined colour (e.g. the same

246   region significant for task version 1(red) and 3(blue) would be coded in purple the

247   same region significant for all task versions would be coded in white). Figure 1C

248   presents the information association maps (z-scores) for all positive associations

249   between information sampling and performance for each condition in turn across the

250   five spatial frequency bands sampled prior to applying the statistical threshold.

251   Importantly, not all task versions produced equally clear profiles of information

252   use. Close inspection of the results reveals that for fear categorizations, it is only in task

253   version 3 – where social interaction and participant engagement are maximised- that

254   both eyes reach significance in the highest spatial frequency band. Similarly, for

255   happiness categorizations it is only in task version 3 that the entire higher spatial

---

[1] Information processing results for the expressions of fear and sadness were not considered further due to the lack of an appropriate comparison baseline and in respect of the relatively small number of trials collected here, which is likely insufficient for a fully stable solution for these more challenging expressions.

[2] Note that our Bubbles solution may be slightly underpowered with only 3,840 trials in total per expression, per task (compared to 16,800 in Smith et al, 2005; or 5,500 in Smith & Merlusca, 2015), which disproportionately impacts the solution in the highest spatial frequency band.

256 frequency mouth reaches significance. Furthermore, when considering the absolute
257 strength of the association between the important pixel locations and performance (the
258 un-thresholded z-scores, presented in Figure 1C) the largest values are generally
259 observed for task version 3, see Table 1[3].
260

	Scale 1	Scale2	Scale3	Scale4	Scale5
Fear V1	4.3	5.3	7.6	**5.9**	3.4
Fear V2	3.8	5.3	7.2	5.6	4.4
Fear V3	**4.6**	**6.3**	**8.5**	5.5	**7.4**
Happy V1	3.6	4.3	5.3	5.7	4.7
Happy V2	3.5	4.0	4.7	**8.8**	**5.3**
Happy V3	**3.8**	**6.1**	**6.6**	4.9	2.5

261

262 *Table 1. Maximal strength of the association between information location and*
263 *performance (measured in z-scores) indicating a stronger association for task version 3*
264 *for scales 1-3 for both fear and happy, and again for scale 5 for fear categorizations.*

265

266 **Motivation Questionnaire results**
267   One participant failed to understand the instructions with regard to the
268 questionnaire (choosing to only answer one of the eight questions at each
269 administration), and the data for one participant, completing one condition, was lost
270 due to experimenter error leaving 28 participants. A one way repeated measures
271 ANOVA, with task version as the factor with 3 levels, was conducted for each question in
272 turn (GG correction reported for violations of sphericity). Significant effects were
273 further explored with post-hoc follow up t-tests (Bonferroni corrected for multiple
274 comparisons), see Figure 2 for average responses per condition. We observed a
275 significant effect of condition for participant's self-reported enjoyment ($F(2,54) = 4.7$,
276 $p=0.013$, $\eta^2 = 0.15$), interest ($F(2,54) = 3.86$, $p=0.027$, $\eta^2=0.13$), desire to do well

---

[3] Note that due to the nature of the paradigm results, information processing results at the individual level are not possible, and therefore only descriptive statistics can be provided.

277   ($F(2,54) = 4.4$, $p=0.016$, $\eta^2=0.14$), pressure felt ($F(1.57,42.27) = 6.6$, $p=0.006$, $\eta^2=0.2$)

278   and connectedness to the experimenter ($F(1.26, 34.06) = 13.2$, $p<0.001$, $\eta^2=0.33$), with a

279   clear trend for an effect also on the effort they expended ($F(2,54) = 3.04$, $p=0.056$,

280     $\eta^2=0.1$). There was no effect of experimental condition on their desire to be good at the

281   task ($F(2,54) = 1.8$, $p=0.17$, $\eta^2=0.06$) or how important they felt the task was ($F(2,54) =$

282   0.74, $p=0.48$, $\eta^2=0.03$).

283   Planned comparisons confirmed that participants enjoyed participating in

284   condition 3 more than condition 2 ($F(1,27) = 6.4$, $p=0.018$, $\eta^2 = 0.19$) but with no such

285   benefit for condition 2 over condition 1 ($F(1,27) = 0.6$, $p=0.45$, $\eta^2=0.02$). Similarly

286   participants expended more effect in condition 3 compared to condition 2 ($F(1,27) =$

287   3.98, $p=0.056$, $\eta^2=0.13$), with no difference between conditions 1 and 2 ($F(1,27) = 0.3$,

288     $p=0.59$. $\eta^2=0.11$). They also tried to do well more for condition 3 than condition 2

289   ($F(1,27) = 4.7$, $p=0.039$, $\eta^2=0.15$) with no difference between conditions 1 and 2

290   ($F(1,27) = 1.35$, $p=0.26$, $\eta^2=0.05$). As expected, participants felt more connected to the

291   experimenter in condition 3 than 2 ($F(1,27) = 15.9$, $p<0.001$, $\eta^2=0.37$), but this came at

292   the cost of feeling more pressure ($F(1,27) = 6.7$, $p =0.013$, $\eta^2= 0.2$). Again there was no

293   difference for either connectedness or pressure felt between conditions 1 and 2

294   ($F(1,27) = 1.3, 0.36$, $p=0.26, 0.55$, $\eta^2=0.05, 0.013$ respectively). Finally, participant's

295   interest in the experiment did not increase significantly between conditions 2 and 3

296   ($F(1,27) = 0.58$, $p=0.45$, $\eta^2=0.02$) but rather there was a trend for interest to be

297   significantly greater for condition 2 than condition 1 ($F(1,27) = 4.0$, $p=0.056$, $\eta^2=0.13$).

298   In an exploratory analysis we then asked whether subjective feelings representing

299   engagement with the task might be directly correlated with markers of task performance

300   (percentage correct, mean number of bubbles) within each task version. We considered self-

301   report measures of effort expended as the best proxy for task engagement and found clear

302   relationships between increased engagement and improvements in the behavioural

303   performance metrics for all task versions, but most so for task version 3 (V1: Accuracy,

304     $r^2(28) = 0.40$, p=0.03, Information required, $r^2(28) = -0.33$, p=0.09; V2: Accuracy, $r^2(28) =$

305   0.52, p=0.005*, Information required: $r^2(28) = -0.46$, p=0.013; V3: Accuracy, $r^2(28) = 0.53$,

306   p=0.004*, Information required: $r^2(28) = -0.49$, p=0.009*, *denotes Bonferoni corrected

307   significant effects). Note that engagement with the task as approximated by effort

308   expended was not directly correlated with 'pressure felt' under any task version ($r^2(28)$

309 = 0.19, 0.03, 0.22, p=0.35,0.86,0.26 respectively) and in particular the increased pressure felt

310 under task version 3 did not seem to be a significant driving force of improved performance

311 (Accuracy, $r^2(28) = 0.089$, Information required, p=0.65, $r^2(28) = -0.16$, p= 0.43). Similarly,

312 increased feelings of connectedness to the researcher did not correlate significantly with

313 performance under any condition (V3: $r^2(28)<0.21$, p>0.28; V1, V2: $r^2(28)<0.23$, p>0.23).

314

315 **Discussion**

316   Here we tested a modified implementation of the Bubbles reverse correlation

317 paradigm that is more appropriate for a developing sample (children) and potentially

318 others for whom the traditional method would make participation very challenging (e.g.

319 individuals with low cognitive ability). Participants completed three versions of the

320 same Bubbles emotion categorization experiment in a single session, with the order of

321 the different versions counter-balanced. With the exception of the reduced number of

322 trials, the first version mirrored that of a standard experiment in most aspects (generic

323 screens providing self-paced short breaks every few minutes, though the use of a count-

324 down bar and presence of an experimenter in the testing room are novel). The second

325 version introduced the puzzle-bubble game as a self-controlled diversion from the

326 monotony of the main task. Finally, in version three, the experimenter actively 'played'

327 the puzzle-bubble game with the participant, acting as quiz master to interact and

328 provide encouragement. Our results indicated better performance for version 3 across

329 the board. Participants demonstrate a capacity to achieve higher performance levels

330 and require less information to do so when performing an otherwise identical

331 psychophysics task. In addition, participants are also subjectively more motivated –

332 they report higher levels of enjoyment, interest, effort and a greater desire to do well.

333 Unsurprisingly, participants also feel a greater connection to the experimenter but also

334 more pressure.

335   A relatively large number of participants for this type of study (thirty) each

336 completed a relatively small number of experimental trials (128 per emotion category)

337 in each of the three different experimental arrangements. Despite a smaller overall

338 number of trials (3840 here (per emotion, per experiment version) vs 5000 [Smith &

339 Merlusca, 2015] or 16800 [Smith, Gosselin, Cottrel & Schyns, 2005]), our Bubbles

340 information use results are clearly aligning with established findings for the well-

341 established happy and fearful expression categorizations. The significant features

342 driving fear categorizations (wide-open eyes across the high and mid spatial

343 frequencies, mouth at lower spatial frequencies) and the features found to be significant

344 for happy categorizations (the broad smiling mouth across spatial frequency bands)

345 mirror past findings. We observe most of these significant visual regions for all three

346 task versions, but note that in fear categorizations the use of the eyes in the highest

347 spatial frequency band only reached significance for version 3. Similarly, it is only under

348 version 3, that the full high spatial frequency mouth reaches significance for happy

349 expression categorization. Furthermore, for the majority of the key visual features, task

350 version 3 produced the statistically cleanest result as indicated by the highest

351 association between visual information and behavioural performance.

352   Our Bubbles paradigm results *and* the motivation questionnaire findings all

353 highlight the importance of social interaction in boosting subjective motivation and task

354 engagement, alongside generating significant improvements in objective task

355 performance and the quality of the Bubbles solution. Little benefit is observed for the

356 use of the game distraction during breaks on its own with the only reported difference

357 being an increase in subjective interest in the task. Past research has shown general

358 cognitive benefits of social interaction including boosting measures of executive

359 functioning (Ybarra et al, 2010), working memory and speed of processing for simple

360 dot patterns (Ybarra et al, 2008) and acting as a potential intervention to slow cognitive

361 decay (Dodge et al, 2015), but to the best of our knowledge this is the first study to find

362 clear benefits of ongoing interaction in a perceptual task such as this. Social interaction

363 is known to consitute a reward in and of itself (Insel, 2003, Walter, Abler, Ciaramidaro &

364 Erk, 2005), and social rewards (typically simply photographs of attractive smiling faces)

365 activate similar neural reward structures as monetary rewards (Aharon et al, 2001;

366 Izuma, Saito & Sadato, 2008; Sprecklemeyer et al, 2009; Lin, Adolphs & Rangel, 2012),

367 with some researchers finding that social rewards can be *even more* motivating than

368 financial rewards in occupation contexts (Graham and Unruh, 1990). The social

369 interaction taking place in version three of the task here could function in a similar, and

370 likely enhanced manner, to activate these same reward structures and boost goal

371 directed behaviour in the task.

372   As such, we conclude that it is likely that any similar diversionary activity that

373 engages the participant with the experimenter during breaks is likely to lead to a

374 similar boost in performance and participant experience. Further studies could explore

375   the extent and nature of the diversion and interaction required in more detail to further

376   optimise testing efficiency. Extant evidence suggests that the interaction should be

377   neutral or cooperative (as opposed to competitive) to drive improved performance

378   (Ybarra et al, 2010). Other factors including explicit feedback, either as vocal praise (a

379   staple of education / training), numerical assessments of ability, or more traditional

380   rewards (e.g. desired foods, monetary rewards, gifts e.g. small toys / stickers for

381   children) could also be interesting avenues to explore in the context of boosting task

382   engagement and associated performance and ability.

383   It is important to note that we did not set out here to establish the necessary (or

384   sufficient) number of trials required to achieve a stable Bubbles solution, and it would

385   be incorrect to conclude a lower bound from the current findings. Determining the

386   necessary number of trials required to accurately categorize important information use

387   for a particular categorization is an important question for future research but is

388   outside the scope of the current manuscript. It is a complex problem that will vary

389   depending on a considerable number of factors. For example, obtaining a stable solution

390   will require more power (i.e. more trials) when the categorization to be made is more

391   challenging, e.g. in the case of sadness and anger here. Trial numbers might also vary if

392   individual differences across participants results in consistently high levels of noise -

393   see Wang, Friel, Gosselin & Schyns (2011) for an estimate in small set of individuals in a

394   standard Bubbles expression categorisation task and note that they observed

395   considerable individual differences in the number of trials required. If it proves possible

396   to establish a target number of trials for a particular categorization, one could then

397   explore if improvements to participant engagement significantly alter this. Finally, it is

398   also important to note that the participant-friendly approach presented here is

399   intended to pull out similarities in information use within a wide sample of participants.

400   In situations where one expects the sample to vary widely in the strategies employed

401   e.g. in developmental prosopagnosics who report a wide number of strategies to

402   counter their face processing deficits (Yardley, McDermott, Poisarski, Duchaine, &

403   Nakayama, 2008), an approach such as this is unlikely to work.

404

405   **Conclusions**

406   Working productively with young children and other groups varying in cognitive

407   ability often requires careful consideration of the participant experience that can be

408   foreign to those working with complex psychophysical paradigms. The results

409   presented here signal that child-friendly design modifications are possible and need not

410   undermine the interpretability of results. In fact, our findings show the opposite

411   pattern. Here, they pay clear research dividends with typical adult participants. By

412   boosting task engagement via an interactive game, we were able to improve objective

413   task performance and the statistical power of our results in a basic investigation of face

414   processing taking place in a short testing session (only 15mins per task). These results

415   will hopefully encourage researchers to see that creating a friendly and engaging

416   participant experience should not be limited to situations with children or atypical

417   populations. We have confirmed empirically that there are significant benefits

418   associated with expending a little more time and effort during data collection.

419

423   **References**

424   Adolphs, R., Gosselin, F., Buchanana, T.W., Tranel, D., Schyns, P. & Damasio, A.R. (2005).

425   A mechanism for impaired fear recognition after amygdala damage. *Nature, 433,* 68–72.

426   Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., Breiter, H.C. (2001). Beautiful

427   faces have variable reward value: fMRI and behavioural evidence. *Neuron,* 32, 537-551.

428   Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the*

429   *Acoustical Society of America*, 49(6), 1751– 1756.

430   Boksem, M. A. S., Meijman, T.F., & Lorist, M. M. (2005). Effects of mental fatigue on

431   attention: An ERP study. *Cognitive Brain Research,* 24, 107-116.

432   Boksem, M. A. S., Meijman, T.F., & Lorist, M. M. (2006). Mental fatigue, motivation and

433   action monitoring. *Biological Psychology,* 72, 123-132.

434   Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and Benefits. *Brain Research*

435   *Reviews,* 59, 125-139.

436   Borghini, G., Astolfi, L. Vecchiato, G., Mattia, D., & Babiloni, F. (2012). Measuring

437   neurophysiological signals in aircraft pilots and car drivers for the assessment of mental

438   workload, fatigue and drowsiness. *Neuroscience Biobehavioral Reviews,* 44, 58-75.

439   Butler, S., Blais, C., Gosselin, F., Bub, D., & Fiset, D. (2010). Recognizing famous people.

440   *Attention, Perception & Psychophysics*, 72, 1444-1449.

441   Caldara, R., Schyns, P.G., Mayer, E., Smith M. L., Gosselin, F., & Rossion, B. (2005). Does

442   prosopagnosia take the eyes out of face representations? Evidence for a defect in

443   representing diagnostic facial information following brain damage. *Journal of Cognitive*

444   *Neuroscience*, 17, 1652-1666.

445   Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate

446   statistical tests for smooth classification images. *Journal of Vision*, 5, 659-667.

447   doi:10.1167/5.9.1

448   Dailey, M., Cottrell, G.W., & Reilly, J. (2001). *California Facial Expressions, CAFE.*

449   Unpublished digital images. University of California, San Diego, Computer Science and

450   Engineering Department, Sam Diego, CA.

451   Dodge, H, H., Zhu, J., Mattek, N. C., Bowman, M., Ybarra, O., Wild, K. V., Loewenstein, D. A.,

452   & Kaye, J. A. (2015). Web-enabled conversational interactions as a method to improve

453   cognitive functions: Results of a 6-week randomized controlled trial. *Alzheimer's &*

454   *Dementia: Translational Research and Clinical Interventions*, 1, 1-12.

455   Ewing et al, 2017a,

456   Ewing et al, 2017b

457   Ewing et al, 2017c

458   Gosselin, F. & Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information

459   in recognition tasks*. Vision Research*, 41, 2261-2271.

460   Haig, N. D. (1985). How faces differ – a new comparative technique. *Perception,* 14(5),

461   601-615.

462   Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?.

463   *Behavioral and Brain Sciences*, *33*(2-3), 61-83.

464   Hopstaken, J. F., van der Linder, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A

465   multifaceted investigation of the link between mental fatigue and task disengagement.

466   *Psychophysiology,* 52, 305-315.

467   Hopstaken, J. F., van der Linder, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The

468   window of my eyes: Task disengagement and mental fatigue covary with pupil

469   dynamics. *Biological Psychology,* 110, 100-106.

470   Insel, T. R. (2003). Is social attachment an addictive disorder? *Physiology & Behavior,* 79,

471   351-357.

472   Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in

473   the human striatum. *Neuron,* 58, 284-294.

474   Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage

475   overlapping neural substrates. *SCAN,* 7, 274-281.

476   Lorist, M. M., Boksen, M. A., & Ridderinkof, K.R. (2005). Impaired cognitive control and

477   reduced circulate activity during mental fatigue. *Brain Research, Cognitive Brain*

478   *Research,* 24, 199-205.

479   Marmarelis, P. Z., & Marmarelis, V. Z. (1978). Analysis of physiological systems: The white-

480   noise approach. New York: Plenum Press.

481   Royer, J., Blais, C., Gosselin, F., Duncan, J. & Fiset, D. (2015). When less is more: The

482   impact of face processing ability on the recognition of visually degraded faces. *Journal of*

483   *Experimental Psychology: Human Perception and Performance*.

484   Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of

485   cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461.

486   Sekuler, A. B., Gaspar, C., Gold, J.M. & Bennett, P. J. (2004). Inversion leads to quantitate,

487   not qualitative, changes in face processing. *Current Biology*, 14, 391-396.

488   Schyns, P.G. & Oliva, A. (1999). Doctor Angry and Mr Smile: when categorization flexibly

489   modifies the perception of faces in rapid visual presentations. *Cognition,* 69(3), 243-

490   265.

491   Smith, M.L., Fries, P., Gosselin, F., Goebel, R., & Schyns, P.G (2009). Inverse Mapping the

492   Neuronal Substrates of Face Categorizations, *Cerebral Cortex*, 19, 2428-2438.

493   Smith, M. L., Gosselin, F. & Schyns, P. G., (2004). Receptive fields for flexible face

494   categorizations. *Psychological Science*, 15, 753-761.

495   Smith, M.L., Volna, B., & Ewing, L. (2016). Distinct information critically distinguishes

496   judgements of face familiarity and identity. *Journal of Experimental Psychology: Human*

497   *Perception and Performance,* 42(11), 1770-1779.

498   Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., Gosselin, F., Hadley,

499   D., Cavanagh, J., & Schyns, P.G. (2008). Classification Images reveal the Information

500   Sensitivity of brain voxels in fMRI. *NeuroImage*, 40, 1643-1654.

501   Smith, F.W., & Schyns, P. G. (2009). Smile through your fear and sadness. *Psychological*

502   *Science,* 20, 1202-1208. doi: 10.1111/j.1467-9280.2009.02427.x

503   Schyns, P.G., Petro, L., & Smith, M.L. (2007). Dynamics of visual information integration

504   in the brain to categorize facial expressions. *Current Biology, 17,* 1580-1585.

505   Schyns, P.G., Gosselin, F., & Smith, M.L. (2009). Information processing algorithms in the

506   brain. *Trends in Cognitive Sciences.*

507   Smith, M.L., Gosselin, F., Cottrell, G.W., & Schyns, P.G. (2005). Transmitting and decoding

508   facial expressions. *Psychological Science, 16,* 184-189.

509   Spreckelmeyer, K., N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., Kircher,

510   T., & Grunder, G. (2009). Anticipations of monetary and social reward differentially

511   activates mesolimbic brain structures in men and women. *SCAN,* 4, 158-165.

512   Tanaka, M., Ishii, A., & Watanabe, Y. (2014). Neural effects of mental fatigue caused by

513   continuous attention load: a magnetoencephalography study. *Brain Research,* 1561, 60-

514   66.

515   van der Linden, D., & Eling, P. (2006). Mental fatigue disturbs local processing more

516   than global processing. *Psychological Research,* 70, 395-402.

517   van der Linder, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of

518   cognitive processes: effects on perseveration and planning. *Acta Psychologica,* 113, 45-

519   65.

520   Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008).

521   Psychosocial consequences of developmental prosopagnosia: A problem of recognition.

522   *Journal of Psychosomatic Research, 65,* 445 – 451.

523   Walter, H., Abler, B., Ciaramidaro, A., & Erk, S. (2005) Motivating foces of human actions:

524   Neuroimaging reward and social interaction. *Brain Research Bulletin,* 67, 368-381.

525   Wang, H. F., Friel, N., Gosselin, F., & Schyns, P. G. (2011). Efficient bubbles for visual

526   categorization tasks, *Vision Research,* 1318-1323.

527   Ybarra, O., Burnstein, E., Winkielman, P., Keller, M. C., Manis, M., Chan, E., & Rodriguez, J.

528   (2008). Mental exercising through simple socialising: social interaction promotes

529   general cognitive functioning. *Personality and Social psychology bulletin,* 34, 248-259.

530   Ybarra, O., Winkielman, P., Yeh, I., Burnstein, E., & Kavanagh, L. (2010). Friends (and

531   sometimes enemies) with cognitive benefits: what types of social interactions boost

532   executive functioning? *Social Psychological and Personality Science,* 1-9.

533

534

535

536

Figure Captions

Figure 1. A – behavioural metrics of performance accuracy and the amount of information required in red for condition 1, green for condition 2 and blue for condition. B – regions significantly associated with correct categorization performance for fearful and happiness categorizations ($p<0.05$ corrected) for condition 1 (red), condition 2 (green) and condition 3 (blue). Note that when the same region is significant for multiple conditions it will be coloured as per the RGB colour space combination (e.g. purple = red + blue = condition 1 and condition 3, white = red+green+blue = all three conditions). C – the un-thresholded information association maps between correct categorization performance and information location (measured as z-scores, higher values represent a greater association between presentation of information at that location and correct categorization response).

Figure 2. Subjective ratings from participants after completing each experimental condition (condition 1 in red, condition 2 in green and condition 3 in blue).
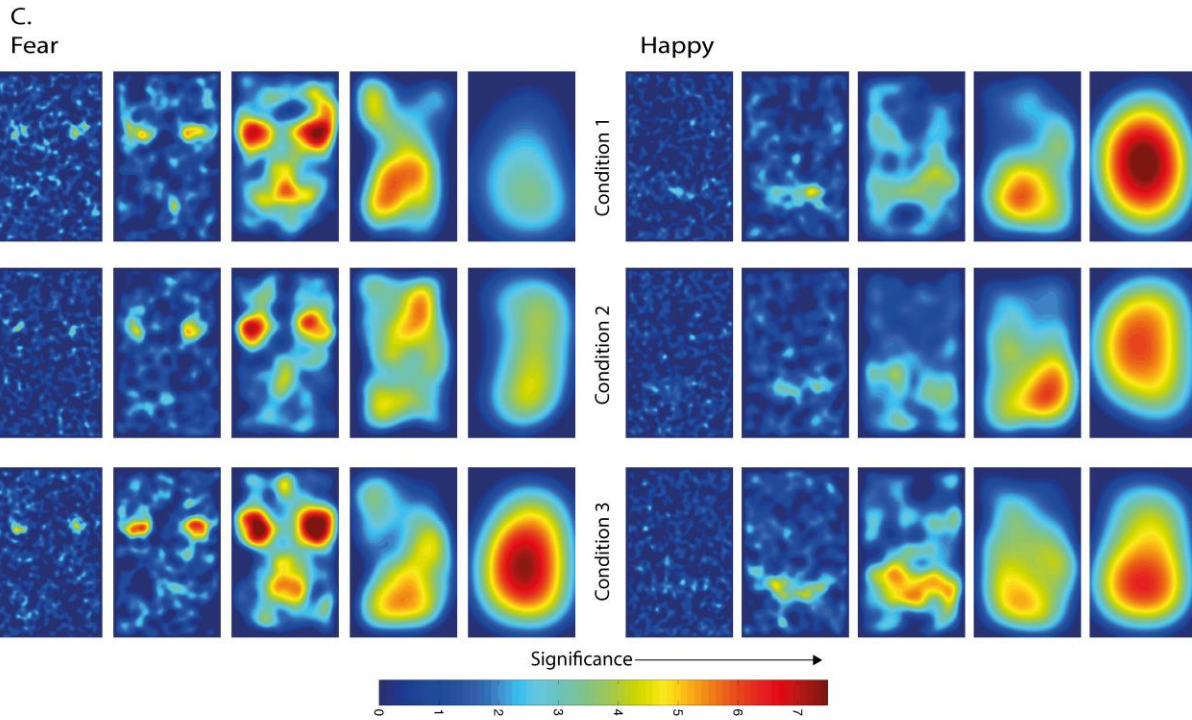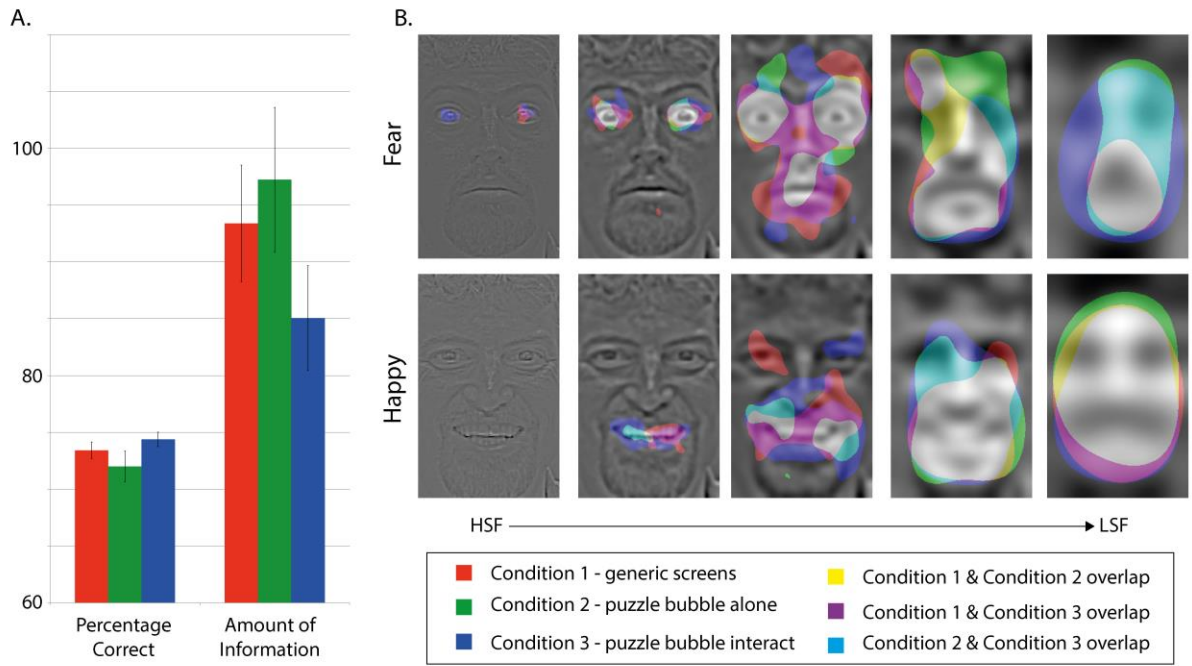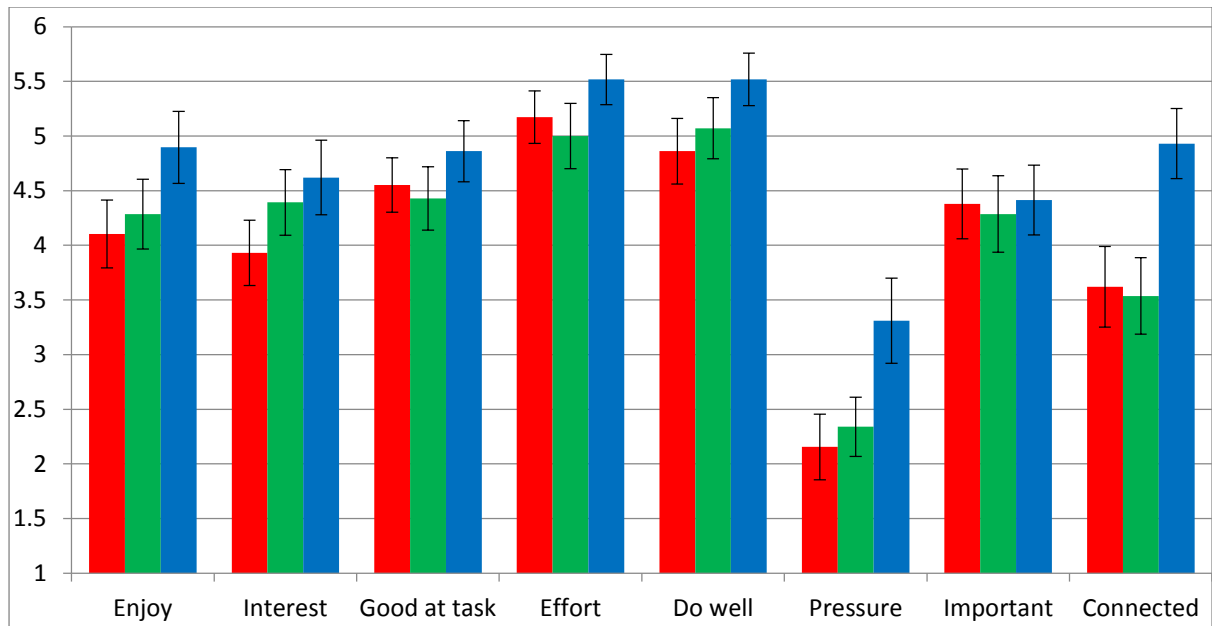
555

556

557

558   Figure 1.

559

560

561   Figure 2.