A 'spoon full of sugar' helps the medicine go down: how a participant friendly version of a psychophysics task significantly improves task engagement, performance and data quality in a typical adult sample.

Marie L. Smith[1], M. Letizia Cesana[1,2], Emily K. Farran[3], Annette Karmiloff-Smith[1] & Louise Ewing[1,4, 5]

[1]Psychological Sciences, Birkbeck College, University of London, Malet Street, London, WC1E 7HX

[2] *Department of Psychology, Catholic University of the Sacred Heart, Largo Gemelli 1, 20123 Milan, Italy.*

[3]Psychology and Human Development, UCL Institute of Education, University College London, 25 Woburn Square, London WC1H 0AA

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

[5] ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, 35 Stirling Hwy, Perth, 6009

Abstract

Few would argue that the unique insights brought by studying the typical and atypical development of psychological processes are essential to building a comprehensive understanding of the brain. Often, however, the associated challenges of working with non-standard adult populations results in the more complex psychophysical paradigms being rejected as too complex. Recently we created a child (and clinical group) friendly implementation of one such technique – the reverse correlation Bubbles approach and noted an associated performance boost in adult participants. Here, we compare the administration of three different versions of this participant-friendly task in the same adult participants to empirically confirm that introducing elements in the experiment with the sole purpose of improving the participant experience, not only boost the participant's engagement and motivation for the task but results in significantly improved objective task performance and stronger statistical results.

Keywords

Reverse correlation, Bubbles, task engagement, child friendly, experimental outcomes

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 breaks, and finally the introduction of the puzzle-bubble game during breaks. This game involved the participants guessing the name of famous films/locations/tv-shows from as little visual information as possible; cheeky bubbles 'hid' the key details, but could be removed by the experimenter to provide further clues. Anecdotally, these changes and the puzzle bubble game in particular, appeared surprisingly popular for children and adults alike!

Although the effects of mental fatigue are well-known to negatively impact cognitive performance (e.g. Boksem, Meijman & Lorist, 2005; Hopstaken, van der Linder, Bakker & Koppier, 2015) with underlying changes in brain activation patterns (e.g. Lorist, Boksen & Ridderinkof, 2005; Boksem, Meijman & Lorist, 2006; Borghini et al, 2012; Tanaka, Ishii & Watanabe, 2014) there tends to be only minimal consideration of the participant experience during the administration of repetitive tasks often asked of participants in Psychology experiments. Mental fatigue occurs as a result of sustained periods of demanding task performance and is typically characterized by changes in mood and motivation (e.g. Boksem & Tops, 2008), and in particular a reduction in task engagement (Hopstaken, van der Linden, Bakker & Kompier, 2015a). Due to its importance in driving workplace errors and accidents, the study of mental fatigue has often focused on the practical implications for occupational settings. However, as mental fatigue is often directly linked to brain processes critical for performance in psychophysical tasks (e.g. attention, Boksem, Meijman, Lorist, 2005, global/local processing bias, van der Linder & Eling, 2006, executive control, van der Linder, Frese, & Meijman, 2003) it follows that by overlooking their impact, researchers of human behaviour may be deleteriously adding noise to their studies.

Research suggests that one way to counter the effects of mental fatigue is to boost the rewards associated with participation (Boksem & Tops, 2008; Hopstaken, van der Linder, Bakker & Kompier, 2015a; Hopstaken, van der Linder, Bakker & Kompier, 2015b) to re-engage fatigued participants in a given task. Given this, we were interested to observe if our participant-friendly task modifications, which were specifically designed to engage young / cognitively impaired individuals in our demanding, repetitive and relatively boring tasks, could also have a measurable impact on task performance and data quality in a standard adult sample.

We set out to validate the impact and effectiveness of our task engagement strategy and the modifications made to the operation of the task by running three identical base versions of the paradigm with the same adult participants in a single testing session. In one version adults performed the task with no experimenter interaction during the entirety of the task. There was no puzzle bubble game and only generic self-paced "take a break" screens between blocks. In a second version, adults again performed the experiment with no experimenter interaction, but with the puzzle bubble game (played independently) separating blocks (even numbered blocks only). Finally, in the third version, the experimenter interacted with the participant as they played the puzzle bubble game (matching the participant-friendly implementation). All other aspects of the methodology remained constant across the three versions of the task. Furthermore, we employed a modified short-form of the Intrinsic Motivation Inventory (IMI, Ryan, 1982) to directly assess each participant's subjective experience of each experimental condition to determine if our manipulations significantly altered the participants' experience of completing the task.

We directly compared performance across the different versions of the task, with the expectation that the introduction of both the puzzle-bubble game (to enforce breaks between trial blocks thus ensuring that the blocks are spaced out and to alleviate the tedium of completing many similar trials) and interaction with the observer during breaks would lead to better performance on the task and cleaner statistical results. Comparing version 2 and version 1 permits us to evaluate the effectiveness of the puzzle bubble game in boosting task engagement in itself, while the comparison of version 2 and version 3 establishes the extent to which any improved performance is driven by interaction with the experimenter. Direct comparison between self-report measures of task engagement (from the IMI) and objective performance metrics (from the Bubbles task) allows us to explicitly establish if greater task engagement is significantly tied to experimental outcomes on a psychophysics task such as this.

To the best of our knowledge, this is the first time that the impact of the participant experience has been explored in the context of a repetitive visual psychophysics task conducted under typical experimental testing conditions (not those designed to specifically induce mental fatigue by having participants perform the same task repeatedly for a number of hours with no breaks). Should the subjective participant experience and task engagement directly impact cognitive performance and resulting data quality, then there are clear implications across a wide range of research areas in the Psychological Sciences.

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 studies, e.g. Smith et al, 2005; Smith & Merluscal, 2014; Schyns, Petro & Smith, 2007). As per existing Bubbles studies of facial expression categorization, stimuli were decomposed in 6 non-overlapping spatial frequency bands (SF) of one octave each (120-60, 60-30, 30-15, 17-7.5, 7.5-3.8 cycles per image). To create a single experimental stimulus, each SF band was sampled independently with randomly positioned Gaussian apertures (the Bubbles) whose size was adjusted at each scale to reveal 3 cycles per aperture and whose number (per scale) was adjusted to ensure equivalent sampling of each SF scale (i.e. more small high SF bubbles than the larger low SF bubbles). The sampled information from each scale was then recombined into a single stimulus image comprising visual information across the SF bands (see Gosselin & Schyns, 2001 and Smith et al, 2005 for fuller details of the stimulus generation process). The total number of apertures (Bubbles) over all SF scales was adjusted on a trial per trial basis via a staircase algorithm to target a performance criterion of 75% correct. To this end, poor performance resulted in more information on a subsequent trial (i.e. more bubbles), while higher than target performance resulted in a reduction in the amount of information presented on subsequent trials (i.e. less bubbles).

In each version of the task participants completed 512 emotion categorization trials (128 per emotion) by categorizing each stimulus by emotion (labelled keyboard keys denoted fearful, happy, angry, sad and don't know), for a total of 1536 trials over the course of the full experiment comprising the three task versions. A short practice phase prior to testing confirmed that participants could correctly categorize the non-Bubbled (i.e., intact) face stimuli by expression and introduced the participants to the response keys. Participants sat 70cm from the experiment which ran on MATLAB using the Psychophysics toolbox (Pelli), such that stimuli subtended a viewing angle of 5.36 x 3.7degrees of visual angle.

Unlike standard implementations of the Bubbles technique, in the modified child-friendly version, we added a 'don't know' response to prevent participants from correctly guessing and adding unnecessary noise to the data (don't know responses were coded as incorrect). Furthermore, we introduced a count-down bar onscreen that permitted participants to gauge their position in a block, and reduced the length of individual blocks to a few minutes (64 trials) rather than around 5minutes.

To gauge interest/motivation, participants completed a short form of the intrinsic motivation inventory assessment (IMI, Ryan, 1982) at the end of each experimental condition. In this questionnaire we asked participants to rate (on a scale of 1 to 7) how they felt about the task they had just completed in terms of their interest and enjoyment (2 separate questions), their perceived competence, the effort they put into their performance, the importance to them of doing well, the degree of pressure they felt, how related they felt to the experimenter and finally how important they felt the task was.

Results

Bubbles results

We considered two performance metrics as dependent measures: the amount of information (i.e. number of bubbles) required to achieve the target performance of 75% correct for each emotion and the actual percentage correct achieved (NB. with a small number of trials it is not possible to perfectly stabilize performance at the target 75% correct), see Figure 1A. Alongside this, we examined the quality of the bubbles solution i.e. the visual information that is significantly associated with categorization of each emotional expression. A one way repeated measures ANOVA with task version (1,2,3) as the within subjects factor indicated a clear main effect on the amount of information required to achieve good performance levels ($F(2,58) = 3.8$, $p=0.029$, $\eta^2=0.12$). Planned comparisons revealed that participants required significantly less information for task version 3 (M = 85 bubbles) compared to task version 2 (M = 97bubbles, $F(1,29) = 5.6$, $p=0.025$, $\eta^2=0.16$), but there was no such drop in number of bubbles for task version 2 compared to task version 1 (M = 93 bubbles, $F(1,29) = 0.9$, $p=0.35$, $\eta^2=0.03$). An equivalent ANOVA on percentage correct scores indicated a trend for a main effect of condition here too ($F(1.3, 37.6) = 3.4$, $p=0.06$, $\eta^2=0.11$), with planned comparisons again showing that participants are performing slightly better in task version 3 (74.4%) compared to task version 2 (72%, $F(1,29) = 4.2$, $p=0.049$, $\eta^2=0.13$), but with no improvement for task version 2 compared to task version 1 (73.4%, $F(1,29) = 2.3$, $p=0.14$, $\eta^2=0.07$).

To evaluate the effectiveness of the task version manipulations on the quality of the bubbles solution we considered the information processing results for the two most well researched emotional expression categorizations: fear and happiness[1]. The critical visual information for both fear and happiness categorizations has been confirmed across a number of studies in typical adult participants. For fearful categorizations the crucial visual information has been repeatedly shown to comprise the wide-open eyes across scales in higher spatial frequencies (scales 1-3), alongside the open mouth (scales 3 and 4, e.g. Smith & Merlusca, 2015; Smith et al, 2005, F. Smith & Schyns, 2009; Adolphs et al, 2005). For happiness categorisations it is the wide-open mouth, from fine detail in the higher spatial frequencies through to the broad low spatial frequency mouth shape information (Smith & Merlusca, 2015; Smith et al, 2005, F. Smith & Schyns, 2009; Adolphs et al, 2005).

For both fear and happy, and all three task version scenarios, the bubbles solution replicates *most*[2] of the key features of these established processing profiles. Figure 1B shows only those regions that pass the corrected statistical tests ($p<0.05$, Chauvin et al, 2005) highlighted on a sample face. Significant regions observed under task version 1 are coded in red, those from task version 2 in green and finally those of task version 3 in blue. Note that where the same regions were significant in multiple task versions it is colour coded in the RGB colour space combined colour (e.g. the same region significant for task version 1(red) and 3(blue) would be coded in purple the same region significant for all task versions would be coded in white). Figure 1C presents the information association maps (z-scores) for all positive associations between information sampling and performance for each condition in turn across the five spatial frequency bands sampled prior to applying the statistical threshold.

Importantly, not all task versions produced equally clear profiles of information use. Close inspection of the results reveals that for fear categorizations, it is only in task version 3 – where social interaction and participant engagement are maximised- that both eyes reach significance in the highest spatial frequency band. Similarly, for 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.

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

	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

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

Motivation Questionnaire results

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

(F(2,54) = 4.4, p=0.016, η^2=0.14), pressure felt (F(1.57,42.27) = 6.6, p=0.006, η^2=0.2) and connectedness to the experimenter (F(1.26, 34.06) = 13.2, p<0.001, η^2=0.33), with a clear trend for an effect also on the effort they expended (F(2,54) = 3.04, p=0.056, η^2=0.1). There was no effect of experimental condition on their desire to be good at the task (F(2,54) = 1.8, p=0.17, η^2=0.06) or how important they felt the task was (F(2,54) = 0.74, p=0.48, η^2=0.03).

Planned comparisons confirmed that participants enjoyed participating in condition 3 more than condition 2 (F(1,27) = 6.4, p=0.018, η^2 = 0.19) but with no such benefit for condition 2 over condition 1 (F(1,27) = 0.6, p=0.45, η^2=0.02). Similarly participants expended more effect in condition 3 compared to condition 2 (F(1,27) = 3.98, p=0.056, η^2=0.13), with no difference between conditions 1 and 2 (F(1,27) = 0.3, p=0.59. η^2=0.11). They also tried to do well more for condition 3 than condition 2 (F(1,27) = 4.7, p=0.039, η^2=0.15) with no difference between conditions 1 and 2 (F(1,27) = 1.35, p=0.26, η^2=0.05). As expected, participants felt more connected to the experimenter in condition 3 than 2 (F(1,27) = 15.9, p<0.001, η^2=0.37), but this came at the cost of feeling more pressure (F(1,27) = 6.7, p =0.013, η^2= 0.2). Again there was no difference for either connectedness or pressure felt between conditions 1 and 2 (F(1,27) = 1.3, 0.36, p=0.26, 0.55 , η^2=0.05, 0.013 respectively). Finally, participant's interest in the experiment did not increase significantly between conditions 2 and 3 (F(1,27) = 0.58, p=0.45, η^2=0.02) but rather there was a trend for interest to be significantly greater for condition 2 than condition 1 (F(1,27) = 4.0, p=0.056, η^2=0.13).

In an exploratory analysis we then asked whether subjective feelings representing engagement with the task might be directly correlated with markers of task performance (percentage correct, mean number of bubbles) within each task version. We considered self-report measures of effort expended as the best proxy for task engagement and found clear relationships between increased engagement and improvements in the behavioural performance metrics for all task versions, but most so for task version 3 (V1: Accuracy, $r^2(28)$ = 0.40, p=0.03, Information required, $r^2(28)$ = -0.33, p=0.09; V2: Accuracy, $r^2(28)$ = 0.52, p=0.005*, Information required: $r^2(28)$ = -0.46, p=0.013; V3: Accuracy, $r^2(28)$ = 0.53, p=0.004*, Information required: $r^2(28)$ = -0.49, p=0.009*, *denotes Bonferoni corrected significant effects). Note that engagement with the task as approximated by effort expended was not directly correlated with 'pressure felt' under any task version ($r^2(28)$

= 0.19, 0.03, 0.22, p=0.35,0.86,0.26 respectively) and in particular the increased pressure felt under task version 3 did not seem to be a significant driving force of improved performance (Accuracy, $r^2(28) = 0.089$, Information required, $p=0.65$, $r^2(28) = -0.16$, $p= 0.43$). Similarly, increased feelings of connectedness to the researcher did not correlate significantly with performance under any condition (V3: $r^2(28)<0.21$, $p>0.28$; V1, V2: $r^2(28)<0.23$, $p>0.23$).

Discussion

Here we tested a modified implementation of the Bubbles reverse correlation paradigm that is more appropriate for a developing sample (children) and potentially others for whom the traditional method would make participation very challenging (e.g. individuals with low cognitive ability). Participants completed three versions of the same Bubbles emotion categorization experiment in a single session, with the order of the different versions counter-balanced. With the exception of the reduced number of trials, the first version mirrored that of a standard experiment in most aspects (generic screens providing self-paced short breaks every few minutes, though the use of a count-down bar and presence of an experimenter in the testing room are novel). The second version introduced the puzzle-bubble game as a self-controlled diversion from the monotony of the main task. Finally, in version three, the experimenter actively ‘played’ the puzzle-bubble game with the participant, acting as quiz master to interact and provide encouragement. Our results indicated better performance for version 3 across the board. Participants demonstrate a capacity to achieve higher performance levels and require less information to do so when performing an otherwise identical psychophysics task. In addition, participants are also subjectively more motivated – they report higher levels of enjoyment, interest, effort and a greater desire to do well. Unsurprisingly, participants also feel a greater connection to the experimenter but also more pressure.

A relatively large number of participants for this type of study (thirty) each completed a relatively small number of experimental trials (128 per emotion category) in each of the three different experimental arrangements. Despite a smaller overall number of trials (3840 here (per emotion, per experiment version) vs 5000 [Smith & Merlusca, 2015] or 16800 [Smith, Gosselin, Cottrel & Schyns, 2005]), our Bubbles information use results are clearly aligning with established findings for the well-established happy and fearful expression categorizations. The significant features driving fear categorizations (wide-open eyes across the high and mid spatial frequencies, mouth at lower spatial frequencies) and the features found to be significant for happy categorizations (the broad smiling mouth across spatial frequency bands) mirror past findings. We observe most of these significant visual regions for all three task versions, but note that in fear categorizations the use of the eyes in the highest spatial frequency band only reached significance for version 3. Similarly, it is only under version 3, that the full high spatial frequency mouth reaches significance for happy expression categorization. Furthermore, for the majority of the key visual features, task version 3 produced the statistically cleanest result as indicated by the highest association between visual information and behavioural performance.

Our Bubbles paradigm results *and* the motivation questionnaire findings all highlight the importance of social interaction in boosting subjective motivation and task engagement, alongside generating significant improvements in objective task performance and the quality of the Bubbles solution. Little benefit is observed for the use of the game distraction during breaks on its own with the only reported difference being an increase in subjective interest in the task. Past research has shown general cognitive benefits of social interaction including boosting measures of executive functioning (Ybarra et al, 2010), working memory and speed of processing for simple dot patterns (Ybarra et al, 2008) and acting as a potential intervention to slow cognitive decay (Dodge et al, 2015), but to the best of our knowledge this is the first study to find clear benefits of ongoing interaction in a perceptual task such as this. Social interaction is known to consitute a reward in and of itself (Insel, 2003, Walter, Abler, Ciaramidaro & Erk, 2005), and social rewards (typically simply photographs of attractive smiling faces) activate similar neural reward structures as monetary rewards (Aharon et al, 2001; Izuma, Saito & Sadato, 2008; Sprecklemeyer et al, 2009; Lin, Adolphs & Rangel, 2012), with some researchers finding that social rewards can be *even more* motivating than financial rewards in occupation contexts (Graham and Unruh, 1990). The social interaction taking place in version three of the task here could function in a similar, and likely enhanced manner, to activate these same reward structures and boost goal directed behaviour in the task.

As such, we conclude that it is likely that any similar diversionary activity that engages the participant with the experimenter during breaks is likely to lead to a similar boost in performance and participant experience. Further studies could explore the extent and nature of the diversion and interaction required in more detail to further optimise testing efficiency. Extant evidence suggests that the interaction should be neutral or cooperative (as opposed to competitive) to drive improved performance (Ybarra et al, 2010). Other factors including explicit feedback, either as vocal praise (a staple of education / training), numerical assessments of ability, or more traditional rewards (e.g. desired foods, monetary rewards, gifts e.g. small toys / stickers for children) could also be interesting avenues to explore in the context of boosting task engagement and associated performance and ability.

It is important to note that we did not set out here to establish the necessary (or sufficient) number of trials required to achieve a stable Bubbles solution, and it would be incorrect to conclude a lower bound from the current findings. Determining the necessary number of trials required to accurately categorize important information use for a particular categorization is an important question for future research but is outside the scope of the current manuscript. It is a complex problem that will vary depending on a considerable number of factors. For example, obtaining a stable solution will require more power (i.e. more trials) when the categorization to be made is more challenging, e.g. in the case of sadness and anger here. Trial numbers might also vary if individual differences across participants results in consistently high levels of noise - see Wang, Friel, Gosselin & Schyns (2011) for an estimate in small set of individuals in a standard Bubbles expression categorisation task and note that they observed considerable individual differences in the number of trials required. If it proves possible to establish a target number of trials for a particular categorization, one could then explore if improvements to participant engagement significantly alter this. Finally, it is also important to note that the participant-friendly approach presented here is intended to pull out similarities in information use within a wide sample of participants. In situations where one expects the sample to vary widely in the strategies employed e.g. in developmental prosopagnosics who report a wide number of strategies to counter their face processing deficits (Yardley, McDermott, Poisarski, Duchaine, & Nakayama, 2008), an approach such as this is unlikely to work.

Conclusions

Working productively with young children and other groups varying in cognitive ability often requires careful consideration of the participant experience that can be foreign to those working with complex psychophysical paradigms. The results presented here signal that child-friendly design modifications are possible and need not undermine the interpretability of results. In fact, our findings show the opposite pattern. Here, they pay clear research dividends with typical adult participants. By boosting task engagement via an interactive game, we were able to improve objective task performance and the statistical power of our results in a basic investigation of face processing taking place in a short testing session (only 15mins per task). These results will hopefully encourage researchers to see that creating a friendly and engaging participant experience should not be limited to situations with children or atypical populations. We have confirmed empirically that there are significant benefits associated with expending a little more time and effort during data collection.

Acknowledgements

This research was supported by a grant from the Leverhulme Trust: RPG-2013-019 awarded to MLS, EF and AKS.

References

Adolphs, R., Gosselin, F., Buchanana, T.W., Tranel, D., Schyns, P. & Damasio, A.R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433,* 68–72.

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., Breiter, H.C. (2001). Beautiful faces have variable reward value: fMRI and behavioural evidence. *Neuron,* 32, 537-551.

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49(6), 1751– 1756.

Boksem, M. A. S., Meijman, T.F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research,* 24, 107-116.

Boksem, M. A. S., Meijman, T.F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology,* 72, 123-132.

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and Benefits. *Brain Research Reviews,* 59, 125-139.

Borghini, G., Astolfi, L. Vecchiato, G., Mattia, D., & Babiloni, F. (2012). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience Biobehavioral Reviews,* 44, 58-75.

Butler, S., Blais, C., Gosselin, F., Bub, D., & Fiset, D. (2010). Recognizing famous people. *Attention, Perception & Psychophysics*, 72, 1444-1449.

Caldara, R., Schyns, P.G., Mayer, E., Smith M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience*, 17, 1652-1666.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5, 659-667. doi:10.1167/5.9.1

Dailey, M., Cottrell, G.W., & Reilly, J. (2001). *California Facial Expressions, CAFE.* Unpublished digital images. University of California, San Diego, Computer Science and Engineering Department, Sam Diego, CA.

Dodge, H, H., Zhu, J., Mattek, N. C., Bowman, M., Ybarra, O., Wild, K. V., Loewenstein, D. A., & Kaye, J. A. (2015). Web-enabled conversational interactions as a method to improve cognitive functions: Results of a 6-week randomized controlled trial. *Alzheimer's & Dementia: Translational Research and Clinical Interventions*, 1, 1-12.

Ewing et al, 2017a,

Ewing et al, 2017b

Ewing et al, 2017c

Gosselin, F. & Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks*. Vision Research*, 41, 2261-2271.

Haig, N. D. (1985). How faces differ – a new comparative technique. *Perception,* 14(5), 601-615.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, *33*(2-3), 61-83.

Hopstaken, J. F., van der Linder, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology,* 52, 305-315.

Hopstaken, J. F., van der Linder, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology,* 110, 100-106.

Insel, T. R. (2003). Is social attachment an addictive disorder? *Physiology & Behavior,* 79, 351-357.

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron,* 58, 284-294.

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *SCAN,* 7, 274-281.

Lorist, M. M., Boksen, M. A., & Ridderinkof, K.R. (2005). Impaired cognitive control and reduced circulate activity during mental fatigue. *Brain Research, Cognitive Brain Research,* 24, 199-205.

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). Analysis of physiological systems: The white-noise approach. New York: Plenum Press.

Royer, J., Blais, C., Gosselin, F., Duncan, J. & Fiset, D. (2015). When less is more: The impact of face processing ability on the recognition of visually degraded faces. *Journal of Experimental Psychology: Human Perception and Performance*.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461.

Sekuler, A. B., Gaspar, C., Gold, J.M. & Bennett, P. J. (2004). Inversion leads to quantitate, not qualitative, changes in face processing. *Current Biology*, 14, 391-396.

Schyns, P.G. & Oliva, A. (1999). Doctor Angry and Mr Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition,* 69(3), 243-265.

Smith, M.L., Fries, P., Gosselin, F., Goebel, R., & Schyns, P.G (2009). Inverse Mapping the Neuronal Substrates of Face Categorizations, *Cerebral Cortex*, 19, 2428-2438.

Smith, M. L., Gosselin, F. & Schyns, P. G., (2004). Receptive fields for flexible face categorizations. *Psychological Science*, 15, 753-761.

Smith, M.L., Volna, B., & Ewing, L. (2016). Distinct information critically distinguishes judgements of face familiarity and identity. *Journal of Experimental Psychology: Human Perception and Performance,* 42(11), 1770-1779.

Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., Gosselin, F., Hadley, D., Cavanagh, J., & Schyns, P.G. (2008). Classification Images reveal the Information Sensitivity of brain voxels in fMRI. *NeuroImage*, 40, 1643-1654.

Smith, F.W., & Schyns, P. G. (2009). Smile through your fear and sadness. *Psychological Science,* 20, 1202-1208. doi: 10.1111/j.1467-9280.2009.02427.x

Schyns, P.G., Petro, L., & Smith, M.L. (2007). Dynamics of visual information integration in the brain to categorize facial expressions. *Current Biology, 17,* 1580-1585.

Schyns, P.G., Gosselin, F., & Smith, M.L. (2009). Information processing algorithms in the brain. *Trends in Cognitive Sciences*.

Smith, M.L., Gosselin, F., Cottrell, G.W., & Schyns, P.G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16,* 184-189.

Spreckelmeyer, K., N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., Kircher, T., & Grunder, G. (2009). Anticipations of monetary and social reward differentially activates mesolimbic brain structures in men and women. *SCAN,* 4, 158-165.

Tanaka, M., Ishii, A., & Watanabe, Y. (2014). Neural effects of mental fatigue caused by continuous attention load: a magnetoencephalography study. *Brain Research,* 1561, 60-66.

van der Linden, D., & Eling, P. (2006). Mental fatigue disturbs local processing more than global processing. *Psychological Research,* 70, 395-402.

van der Linder, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica,* 113, 45-65.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research, 65,* 445 – 451.

Walter, H., Abler, B., Ciaramidaro, A., & Erk, S. (2005) Motivating foces of human actions: Neuroimaging reward and social interaction. *Brain Research Bulletin,* 67, 368-381.

Wang, H. F., Friel, N., Gosselin, F., & Schyns, P. G. (2011). Efficient bubbles for visual categorization tasks, *Vision Research,* 1318-1323.

Ybarra, O., Burnstein, E., Winkielman, P., Keller, M. C., Manis, M., Chan, E., & Rodriguez, J. (2008). Mental exercising through simple socialising: social interaction promotes general cognitive functioning. *Personality and Social psychology bulletin,* 34, 248-259.

Ybarra, O., Winkielman, P., Yeh, I., Burnstein, E., & Kavanagh, L. (2010). Friends (and sometimes enemies) with cognitive benefits: what types of social interactions boost executive functioning? *Social Psychological and Personality Science,* 1-9.

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).

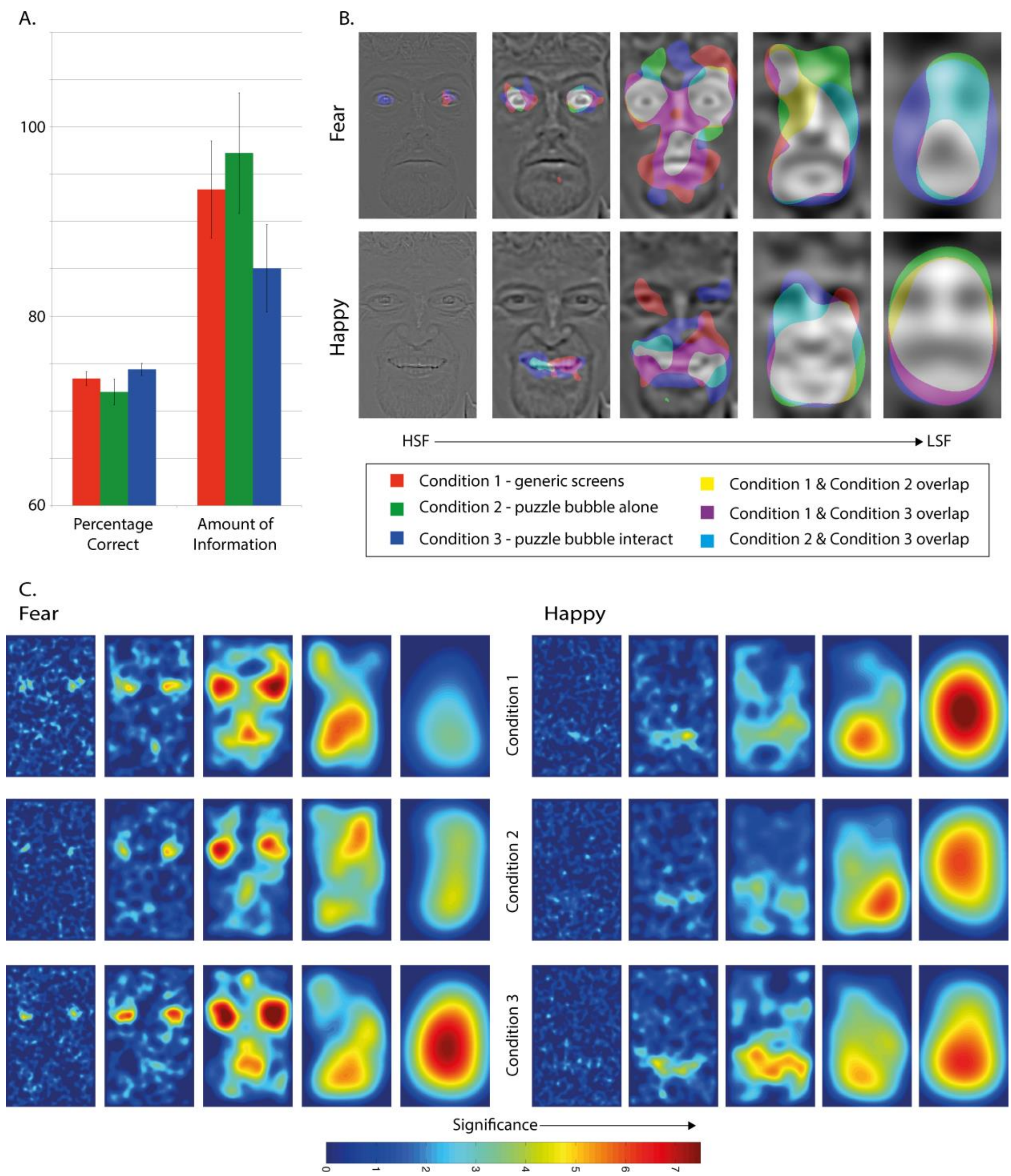
A.
100
80
60
Percentage Correct
Amount of Information
B.
Fear
Happy
HSF
LSF
Condition 1 - generic screens
Condition 2 - puzzle bubble alone
Condition 3 - puzzle bubble interact
Condition 1 & Condition 2 overlap
Condition 1 & Condition 3 overlap
Condition 2 & Condition 3 overlap
C.
Fear
Happy
Condition 1
Condition 2
Condition 3
Significance
0
1
2
3
4
5
6
7

Figure 1.

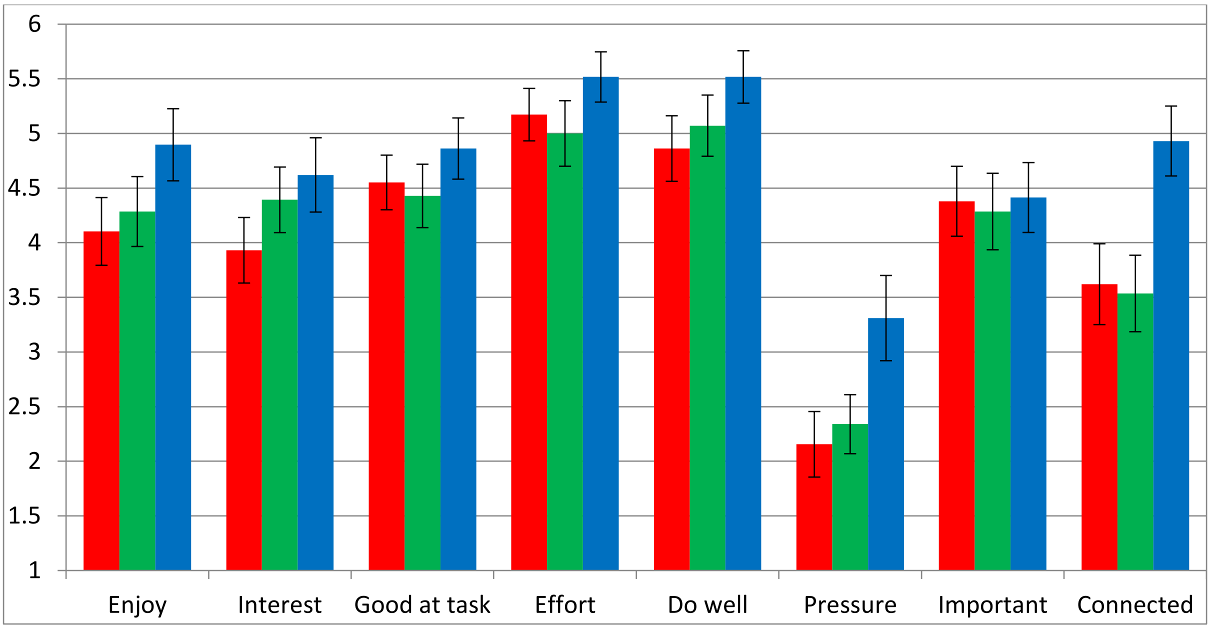
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
Enjoy
Interest
Good at task
Effort
Do well
Pressure
Important
Connected

Figure 2.