Physical Sciences Ethics Committee Full Application Form

Section 1: PROJECT INFORMATION

1.1 Name and email of Principal Investigator (PI):

David Gundry, deg500@york.ac.uk

1.2 Role of PI (Researcher / lecturer / Student):

PhD Student, Department of Computer Science

1.3 Names and appointments of additional investigators

Sebastian Deterding, Reader in the Department of Theatre, Film and Television

1.4 Title of the Project:

Enjoyment and data quality in a human-subject data collection game, a conceptual replication

1.5 Funding source:

The EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/L015846/1] and the Digital Creativity Labs (digitalcreativity.ac.uk), jointly funded by EPSRC/ AHRC/Innovate UK under grant no. EP/M023265/1.

1.6 Where is the research taking place?

Online

1.7 Why are you applying for ethical approval (you must tick a box)?

☐ [X] **It involves people**

☐ [X] **It involves analysing unpublished data from or about living human beings?** (yes, but the data will be subsequently published)

☐ It involves animals?

☐ [X] **It involves data protection**

☐ It is defence / military related

☐ There is a reputational risk to the University

☐ It restricts academic freedom

☐ It involves collaboration / partnership / funding from organisations
tainted by ethically questionable activities?

☐ Other (please state)

_____

1.8 Brief summary of the project and its key aims:

This research is about meta-methodological issues regarding data collection with applied games. In this context an applied game is a game that is designed specifically for collecting a certain type of scientific data.

An applied game has been developed to evaluate how game design affects the validity (accuracy) of data that is collected. In this experiment we do a conceptual replication of a study into the effects of this game on enjoyment and data quality.

A theoretical model and set of design principles for the design of games for subjective data collection has been proposed (http://eprints.whiterose.ac.uk/133118/). This work is the second in a series of experiments to evaluate this model.

The game used is a casual puzzle game. It requires players to identify shapes by clicking on words (adjectives and nouns) to form a 'sentence' of three words.

1.9 External Ethics Approval:
Has this work been approved by another external agency (collaborating institution, NHS) etc.

~~Yes~~ / **No**

If yes please state where:

1.10 I have read and understood the University's Code of practice and principles for good ethical governance (https://www.york.ac.uk/staff/research/governance/research-policies/ethics-code/)

Yes / ~~No~~

Section 2: HUMAN PARTICIPANTS

(If your project doesn't involve people please skip to section 3)

- *You must include an information sheet and consent form alongside this application form (or give a reason that they aren't needed).*

Detailed study and consent information will be included in the study description on Prolific. By clicking to begin the study, they indicate their consent. The description provided to participants is supplied in the attached file `description.md`.

Participants on Prolific are able to choose whether or not to participate in a study based on the study's title and description as well as the expected time taken, payment offered, and number of participant slots remaining. Before choosing to participate, they are shown the study description, which should give all the information needed in a consent form/information sheet. This is to avoid wasting participant time as they might otherwise drop out after first seeing this information within the study questionnaire/task itself. It also avoids having to refuse payment for as many incomplete sessions.

A participant that is delivered to the study by Prolific (as opposed to for some reason stumbling upon the direct url) is associated with a Prolific participant ID (supplied via a GET parameter). Thus if a participant has such a number, we know they have come from Prolific and chosen there to begin the study. In the unlikely event that a participant arrives without such a number their data would be destroyed.

The cost of running the study on Prolific is based on the average completion time of the task. While I have over-estimated this to avoid participants rushing, in general I want to minimise the time the participant spends on the study: both to minimise cost, and avoid wasting the participants' time. This is why I want to avoid duplicating the study information and consent at the beginning of the study. It would be more screens of text to read through.

In other words, I believe participants have supplied informed consent by reading the description and choosing to begin the study. I have an indirect record of this (in the form of a participant arriving to the study with a prolific participant ID). Specifically with regard to the GDPR, under the "public task" justification of the GDPR, a record of consent is not necessary for data processing (which is itself only applicable until it is anonymised and thus ceases to be personal data under the definition of the GDPR).

- *If a survey / questionnaire is being conducted please include this alongside the application.*

- *If you are conducting an online survey these can be embedded at the front of the survey and need to be included in this application.*

2.1 Who will your participants be? (Describe the criteria for inclusion / exclusion)

Participants will be users of Prolific (www.prolific.co), a panel provider for research studies.

- To be registed on Prolific, participants must agree to terms and conditions which include
  - that they are at least 18 years old;
  - that they have authority to enter into legally binding contracts (i.e. they have capacity)
- Participants will be pre-filtered to include only adults (18+) whose first langauge is English.

2.2 Will they be paid? **Yes** / ~~No~~

If yes how much (you must obtain a signed receipt of payment)?

£1. Payment is handled and recorded through Prolific.

2.3 Do any of the following apply?

Children (under 18)
~~Yes~~ / **No**

If yes does the investigator have a current DBS check? Yes / No

Vulnerable groups
~~Yes~~ / **No**

The research is designed to be emotive or aversive
~~Yes~~ / **No**

It involves taking bodily samples
~~Yes~~ / **No**

Is physically invasive / challenging ~~Yes~~ / **No**

If you answered yes to any of the above explain and justify the procedure and explain the steps taken to safeguard individuals:

2.4 Recruitment (How will you recruit participants?)

A study description will be created on Prolific. The study description will be available for users of Prolific matching the inclusion criteria to view. Participants may click a button to begin the study if they choose. If they do so their web browser will be directed to the webserver hosting the study.

2.5 How will you guarantee anonymity? (This includes IP addresses and any identifying information)
If anonymity will not be provided explain why this is necessary.

Participants will be for **all practical purposes completely anonymous** at the point of collection.

The data being collected as part of the study is not enough to identify a participant by itself. Personal data consists of age, gender (limited to male, female, other, prefer not to say), and gaming frequency, and that participants speak English as a first langauge. This is not enough data to identify an individual, even if you assume they are a UK resident with a computer and internet connection. It is implausible that the other data collected (moves made in a puzzle game) could be identifying.

However, there exist other data sets, which in combination with the data collected, could theoretically be used to de-anonymise participants. Thus the data as collected might still count as personal data under the GDPR.

There are two data sets that could plausibly be combined in such a way. First, in order to provide its services to reserachers and participants, Prolific collects data about what studies users perform including time of starting and completion. Prolific also identifies users by a pseudonymous identifier, which is recorded by the study for the purposes of managing payments. By matching participant identifiers to Prolific's confidential database, or by similarly comparing start/end times or durations, it would be possible to identify who in Prolific's database had provided what data.

Secondly, the website is hosted by Netlify (https://www.netlify.com/). They, or someone else such as the participant's ISP may be able to record the IP address of the participant with a time of access to the site. In combination with the start time data recorded by the study, this could be used to identify a participant.

2.6 What types of personal data will you collect?

(https://www.york.ac.uk/records-management/dp/)

___ Physiological

___ Video footage / photographs

___ Audio (conversations, spoken tasks etc.)

___ Medical (in which case you are likely to need NHS approval)

___ Personal (names, contact details etc.)

___ Financial

[X] **Other** (Please state)____

Personal data under the GDPR is all data collected that can be associated with an individual. (Though as described above this is only possible in combination with other data sets.)

Using this definition the personal data collected is therefore: age, gender, gaming frequency, Prolific ID, time study starts/ends, moves made in the puzzle game, an enjoyment questionnaire, and grammaticality judgements of English adjective word order.

How will you protect this data?

Data will be submitted encrypted using SSL to a Restdb.io database (https://restdb.io). Restdb is located in Norway, within the jurisdiction of the GDPR. They are granted only the following rights over the content posted to them, emphasis mine:

> By posting Content to the Service, you grant us the right and license to use, modify, publicly perform, publicly display, reproduce, and distribute such Content on and through the Service, **but only for the purpose of delivering the Content on your behalf**. You retain any and all of your rights to any Content you submit, post or display on or through the Service and you are responsible for protecting those rights.

The only way for them to deliver content on my behalf is to someone authenticating with my user account (they also can provide other services which I am not using). When downloading the data, SSL encryption is again used. The data in the database is stored in plaintext. It will be deleted from the database once it is downloaded.

The data will be downloaded to the University computer network filestore and anonymised there.

The data will be fully anonymised as soon as possible after it is downloaded. This will be done by removing participant's Prolific ID, and disassociating duration and gender from the rest of the data. The non anonymised data will then be deleted. It will then be impossible **even with Prolific's or another data set** to identify individuals. This data will be published for research transparency on the Open Science Framework (https://osf.io)

Section 3: DATA STORAGE AND TRANSMISSION

3.1 I have read and understood the University of York's Data Protection Policy (https://www.york.ac.uk/records-management/dp/policy/)

Yes / ~~No~~

3.2 I will keep any data appropriately secure (e.g. in a locked cabinet), maintaining confidentiality and anonymity (e.g. identifiers will be encoded and the code available on a need to know basis) where possible.

Yes / ~~No~~

3.3 Please describe the special precautions will you take to ensure anonymity when linking identifiable data to experimental data:

I will not collect any identifiable data. For more details, see the description on the previous page.

3.4 If your data can be traced to identifiable participants/computer/address:

a) who will be able to access your data?

David (but again, see the description on the previous page)

b) approximately how long will you need to keep it in this identifiable format?

Less than 24 hours.

3.5 If your project requires deviation from traditional data protection practices in research, or raises particular data protection issue please explain here:

3.6 STUDENTS ONLY: Will any identifying data be kept securely by supervisors?

Yes / ~~No~~

If No state why:
No identifiable data will be collected (see description on previous page).

Section 4: RISK ASSESSMENT

4.1 Has a departmental risk assessment been completed for this project, if appropriate?

~~Yes / No /~~ **Not needed**

If no why not?

Section 5: ACADEMIC FREEDOM

5.1 Is there a secrecy clause to the research? ~~Yes~~ / **No**

If yes give details:

Section 6: REPUTATIONAL RISK *(if associated with a collaborative partner see section 7)*

6.1 Why is it appropriate for the University to be associated with this project? *(please also state what action has been taken to mitigate against potential reputational risk)*

Section 7: COLLABORATION WITH QUESTIONABLE ETHICAL STANDARDS / ACTIVITIES

7.1 Explain the nature of the collaboration / partnership and what about the organisation is ethically questionable:

7.2 Why is it appropriate for the University to be associated with this organisation?
(please also state what action has been taken to mitigate against potential reputational risk)

Section 8: COMPLETION

The project team have read and understood this application:

Signed (PI): David Gundry Date: 23/07/2020

Additional Information
You may find the following codes of ethical practice and conduct relevant to your project:

British Psychological Society code of conduct:
http://www.bps.org.uk/the-society/code-of-conduct/code-ofconduct_home.cfm
IEEE
http://www.ieee.org/about/corporate/governance/p7-8.html
IET
http://www.theiet.org/membership/career/ethics/

Royal Academy of Engineering
http://www.raeng.org.uk/policy/engineering-ethics/ethics

The Royal Society
https://royalsociety.org/topics-policy/ethics-conduct/