TMS@40 Hackathon Briefing

The primary objective of **TMS@40** is to write a paper. The provisional title of this paper is:

"The experimental psychology of transcranial magnetic stimulation: A consensus paper"

Why do we need a consensus paper on the experimental psychology of TMS?

- There are many existing consensus & guidance papers on TMS. Most are from scientists with neurology, neurophysiology & medical backgrounds
- The main journals that specialise in TMS – *Clinical Neurophysiology* & *Brain Stimulation* – again, tend to be edited by & publish work from medical researchers
- By contrast, experimental psychology has had *relatively* little input or influence on how TMS studies are conducted (but see, e.g., Bergmann & Hartwigsen, 2021)
- Since TMS is mostly done on awake, behaving participants, experimental psychology has much to offer. (TMS@40 should, **not**, therefore focus on neurophysiology or EMG.)

What is experimental psychology?

Experimental lab-based methods have long been used to understand relationships between stimuli, (neuro)psychological processes & responses in humans & other animals. Arguably, this began with Weber & Fechner's psychophysics in the mid-1800s. Experimental psychology has developed methods to study sensation, perception, attention, associative learning, memory, decision-making, response conflict & more. It has been an extremely successful paradigm.

TMS is a stimulus like any other, so it follows that the methodological rigour of experimental psychology can be brought to bear on understanding how TMS affects the brain & behaviour. For example, stimulus timing, intensity & location may significantly affect cognition & movement.

How can experimental psychology inform TMS research?

Well, that is what you need to decide & write about. You should use any & all of your training so far, whether it is in psychology, physiology, neuroscience, biology, statistics or other allied disciplines, to decide. Below are some general examples of what you might discuss. There is a more specific briefing providing some ideas to get started in each hackathon topic, below.

Experimental design: How should TMS experiments be designed to minimise bias, remove confounds, ensure counterbalancing & randomisation, prevent order- & learning-effects?

Analysis: How should data from TMS experiments be analysed to minimise bias & confounds, to ensure all relevant variance is modelled & to account for individual differences between people?

Reproducibility: After ~175 years of experimental psychology, what kind of methods work & what don't? What lessons apply to the (recent) history of TMS research? What should we *do*?

TMS@40 Hackathon Briefing: General guidance

Edit one GitHub page only!

- Only edit the page your group is allocated to work on (check Hackathon schedule)
- Nominate one person only to edit the page on their device
- Save your edits frequently (e.g., every minute, every paragraph, every point)

Time is short!

- Take a few minutes to read through the guidance notes (where relevant or necessary)
- Quickly decide which issue(s) or topic(s) to work on; the briefings are *suggestions* only!
- Delegate tasks to different members of the group (e.g., looking up references; checking other wiki pages; posting questions or discussions on Slack; drafting or editing text)
- Write quickly now; edit slowly later (today's notes will help tomorrow's paper writing)

Notes & bullets are great!

- You will not have time to write perfect prose on GitHub. So don't try :-)
- Think of the task as taking 'minutes' on your meeting, discussions & decisions
- Summarise issues with brief notes & bullets. Give brief pros & cons. Link to resources

Say whether the group agrees or not!

- For all (major) topics discussed, note whether there is full agreement (6/6), a majority (4 or 5), or a minority (2 or 3) who agree with the statement or proposal
- If there are a range of views and no consensus, note them down briefly

---

Specific tasks for each session:

Hackathon 1 (11:00-11:45)

- Pick one or more topics, & get started quickly (see above)

Hackathon 2 ***(12:00-12:45)***

- Briefly review work done in Hackathon 1; make only minor revisions, if needed
- If there are major disagreements, make notes on what these are
- Pick one or two NEW topics, then get started quickly (see above)

Hackathon 3 ***(14:00-14:45)***

- Briefly review work done in Hackathons 1 & 2; make only minor revisions, if needed
- If there are major disagreements, make notes on what these are
- Pick at least one NEW topic, then get started quickly (see above)

Hackathon 4 ***(15:00-15:45)***

- Briefly review work done in Hackathons 1, 2 & 3; make only minor revisions, if needed
- Summarise everything decided so far; note any remaining major disagreements
- **Choose ONE controversial topic** that is important enough to go to a VOTE
- Add this to a poll on GitHub (via *Discussions*) – **make sure this is ready for 15:45h**

TMS@40 Hackathon Briefing: Experimental Design 1

How should TMS experiments be designed to minimise bias, remove confounds & ensure that we answer the questions we want to answer? What constraints apply to TMS experimental design? What are the trade-offs, pros & cons of each approach?

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Experimental_design_1

General design choices

Theoretical

- What kinds of question are we trying to answer?
- How do we choose a protocol that best answers our question?
- How do we estimate the statistical power of TMS studies; do we need large(r) samples?

TMS type

- Offline versus online: what questions can they answer? What are the pros & cons?
- Single pulse versus bursts: what questions can they answer? What are the pros & cons?

TMS timing

- When does a TMS pulse interfere with brain processing? Can this be before a stimulus is processed by a brain area? Or must it be during processing? Or can it even be after?
- When does a TMS pulse interfere with a brain area's response – before, during or after?
- Should we start by piloting a range of TMS timings ('chronometric' TMS), or are there principled ways to choose a time-point in advance?

Specific design choices

What should we be controlling?

- Participant body, head, arm, & eye position & activity? Should they be relaxed or natural?
- Should we control participant attention with fixation or a (distractor) task?

Primary motor cortex (M1) versus the rest of the brain

- The majority (or largest minority) of TMS has been done on motor cortex. But the majority of experimental psychology studies are *not* motor. What do we learn from motor TMS?
- How do we generalize M1 protocols to other brain areas?

Electromyography (EMG) versus no EMG

- The majority of experimental psychology TMS studies do *not* measure EMG, so how should we adapt the methods & guidance from TMS studies that measure EMG?
- How should we set stimulus intensity in the absence of EMG? Should we set it relative to motor thresholds like a visible twitch, or at a fixed intensity, or something else?

Magnetic resonance imaging (MRI) versus no MRI

- The majority of experimental psychology TMS studies do *not* use MRI scans, so how do we adapt the methods to follow guidance from TMS studies that do use MRI?
- e-Field modelling without individual MRI scans - is it useful in experimental psychology?

TMS@40 Hackathon Briefing: Experimental Design 2

How should TMS experiments be designed when awake, behaving humans are participating? How do we remove bias & confounds, ensure counterbalancing & randomisation, prevent order- & learning-effects? What is the 'social psychology of the TMS experiment' (Orne, 1962)?

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Experimental_design_2

TMS parameters

- How do different TMS timings relative to a stimulus or response change participants' reactions to those stimuli? Does TMS shorten or lengthen reaction times? Do participants wait until the last TMS pulse in a train before responding?
- Do participants try to predict the TMS? How do we set the intervals between TMS pulses to minimise participants' prediction of or preparation for the next TMS pulse?
- When can we use block designs & when must the conditions of interest be randomised?
- What kinds of TMS work on individual trials? Do some TMS effects build up over time?
- How does TMS change the behaviour we want to measure (ie, regardless of brain area)?
- How does TMS intensity &/or randomness affect our studies or participants' reactions?

Control conditions

- What counts as an adequate control for a TMS experiment? What are we controlling?
- What does a control site control for?
- What can control conditions control for? What kind of conditions are needed?
- What kinds of 'sham' condition are there? What do they control for? How do we know?
- Should we be single-blinding our participants? How do we ensure single-blinding?
- Should we be double-blinding our experimenters? Is it possible & how do we ensure it?
- How do we minimise 'demand characteristics'? (i.e., when our participants 'play the role' of a participant & adjust their behaviours according to what they think we want to see)

Participant (dis)comfort

- How do we ask participants to relax? How do these instructions affect the experiment?
- How do we measure participants' perception of the TMS (annoyance, distraction)?
- How does participant perception differ for online versus offline, or other TMS protocols?
- What should we do (or not do) to minimise discomfort? Decrease intensity? Move coil?
- How often should we be providing participants with breaks from tasks or stimulation?
- At what level of discomfort should TMS studies on healthy participants not continue?

Debriefing

- What should we regularly ask our participants about after a TMS study?
- Should we ask about their tolerance for the stimulation? How? A standard questionnaire?
- If we used a sham (or multiple TMS sites), did they detect the sham? Did blinding work?

TMS@40 Hackathon Briefing: Analysis 1

How should data from TMS experiments be ***pre-processed*** to minimise bias & confounds, & to ensure all relevant variance is modelled?

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Analysis_1

Pre-processing is the filtering, de-noising, 'cleaning' & other processing that might be done to a dataset *before* any statistical analysis. It is often not fully-described in scientific papers, but may have substantial consequences for the statistical analysis stage & the conclusions we can draw.

(Please avoid focusing on EMG data & MEP analysis – this is covered by other papers & research fields)

Measures

- What do we need to record or measure during the experiment to help pre-processing?
- Do we need trial-by-trial TMS intensity, coil position & angles? Or anything else?

Modelling what went wrong during the experiment?

- If we have coil position or angle measures (whether robot- or manually-guided), do we use them as a covariate? At which level of the analysis? Or, do we set a fixed threshold for a 'good' trial - e.g., remove data when the coil is >5mm or >10degrees from target?
- If TMS intensity was increased or decreased (e.g., because of discomfort or missing responses), should we delete the data or include trial-by-trial level covariates?

Temporal filtering

- Do we expect TMS responses to change over time, for example with arousal, alertness, learning, association or plasticity? If yes, (how) do we model it?
- If temporal changes dues to TMS are of no theoretical interest, should we filter responses (such as RTs or other decisions), to remove these trends?

'Outlier' removal

- What is an 'outlier' in TMS studies & why?
- When is it acceptable to delete data?
- What evidence & measurements do we need to justify deleting a data-point?
- How should we document these data removals & their consequences on our analyses?

'Normalisation' of data

- When should we 'normalise' or 'standardise' data? (e.g., ratios of baseline measures per participant, or using the standard normal distribution to convert raw data to Z-scores)
- What are the consequences of such normalisation? (e.g., for linear/non-linear effects, or for understanding between-participant variability)
- Should we always present both normalised & raw data analysis in our papers?

Transformation of data

- Researchers often desire 'normally-distributed' data. If we find non-normal (e.g., logarithmic, exponential, skewed, or other) data, what should we do?
- At what level of analysis should any transformation be applied? (raw data per participant, summary data across participants, or (as the statistics assume) only the residuals after the full model is fit – if the latter, this is something discussed by the ***Analysis 2*** guys***...***)

TMS@40 Hackathon Briefing: Analysis 2

How should data from TMS experiments be ***processed*** to minimise bias & confounds, to ensure all relevant variance & individual differences are modelled?

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Analysis_2

Statistical analysis (post-processing) is the final stage of analysis used to draw conclusions about the manipulations, sample &/or population under study. It is usually the main focus of scientific reports.

(Please avoid focusing on EMG data & MEP analysis – this is covered by other papers & research fields)

Individual differences

- How do we distinguish between individual differences vs. just random variation?
- What covariates should we include in our analyses / what should we be measuring?
  - sham effectiveness/detection
  - participant annoyance/discomfort/tolerance
  - e-field model parameters
  - TMS intensity; motor thresholds
  - fatigue, time-of-day, sleep status
  - caffeine or other recreational substances
  - experience of the subject in prior TMS experiments
  - biometrics such as height, weight, age, sex, handedness
- Which of the above might (really, seriously, honestly!) explain between-study differences (e.g., in a meta-analysis, what between-participants measures would you *really* want)?

Capturing changes in TMS variables over time

- How long do we expect TMS effects to last? (think about different TMS protocols)
- How do participants change the way they respond to TMS over time [seconds... years]?
- Do participants 'learn' to become good TMS participants? How do we measure that?
- Are there 'washout' periods after TMS? What are the minimum intervals we should use?

Statistics & modelling

- With ratio measures (e.g., conditioned / unconditioned designs), how do we account for the consequences of changing baselines or denominators?
- How do we account for the dynamic range of responses? (e.g., where on an input-output or psychophysical curve a stimulation or response or participant is)?
- Should we all be moving beyond standard methods like: pre/post subtraction, ANOVA?
- Do we all need to embrace multi-level modelling, unbalanced designs, nested designs, trial-by-trial analyses? Do we all need to use R now, or can any approach work?

Moving towards fully-open data & code sharing

- What are the barriers to sharing our data & code more freely?
- What first step(s) can we make to improve our analysis pipelines?

TMS@40 Hackathon Briefing: Reproducibility 1

After ~175 years of experimental psychology, what kinds of methods work & what don't? What lessons can be applied to the recent history of TMS research? This session focusses on ***within-person & within-lab reproducibility***.

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Reproducibility_1

What can (or do) you do to improve within-person reproducibility?

- Are the same people tested multiple times? How many times? When is something consistent enough to be called a stable 'trait'?
- Are we comfortable with 'responders' vs. 'non-responders'? How is this quantified?
- If we want to find changes from baseline, how many baseline measures do we need?
- Are datasets split into two (discovery & test), e.g., during code or experiment development?
- How do we check the *robustness* of our methods (e.g., to minor differences)?
- What specific level(s) of within-person test-retest reliability are required? (i.e., what does the correlation coefficient need to be before we can say that a TMS method is reliable?)
- What do we do if test-retest reliability is too low for one or more of our measures?

What can (or do) you do to improve within-lab reproducibility?

- Code review – what are the barriers? How many labs (or mult-labs) can do this?
- Split datasets into two (discovery & test)?
- Multiple analysers?
- Multiverse analysis? (i.e., for a given question, use many or all analytic approaches)
- Red team vs. Blue team? Adversarial collaboration? (better within- or between labs?)
- Replicate all experiments at least once before publication?
- What skills (coding, data analysis, stats), do you need to improve reproducibility?

What is the role of pre-registration?

- Does TMS do this (enough)?
- Should we be sharing experimental plans **before** running the study?
- Is TMSMultiLab a good place for sharing pre-experiment plans? If not, then where?
- Will pre-registration benefit your career (because it front-loads the work of getting a paper accepted), or slow it down (because you can't collect data until Year 2, or later?

What are the challenges of reproducibility?

- Limited resources such as time, money or participants?
- Journals, editors or conferences gate-keeping on 'reproducibility' work?
- Degree regulations or employer's or grant or funding restrictions?
- Perception of 'reproducibility work' by others? Who are they & why is it not valued?

TMS@40 Hackathon Briefing: Reproducibility 2

After ~175 years of experimental psychology, what kinds of methods work & what don't? What lessons can be applied to the more recent history of TMS research? This session focusses on ***between-lab reproducibility***.

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Reproducibility_2

What can (or do) you do to improve between-lab reproducibility?

- Meta-analysis & systematic review before experiments? Do you try to read *everything*?
- Share code, meta-data, data & materials?
- Use shared generic multi-lab protocols (ethics, safety, screening, debrief)?
- Use shared specific multi-lab protocols (software, hardware, configurations)?
- What specific methods differ between labs?
  - electrode or stimulator placement
  - 'hotspotting' & thresholding methods
  - heuristics around setting TMS intensity, position or other parameters
  - different approaches to removing ('noise', 'outliers') or selectively-reporting data
  - cult-like behaviours or heuristics in different labs: "*we always do visual inspection of raw data*" versus "*we never do visual inspection of raw data*"

How to work out what matters?

- Create a list of everything that *might* matter, see who agrees or what data exists already?
- Note things that seem to come up regularly in discussion or papers: time-of-day effects; hormone fluctuations; instructions given; posture/position; lighting…
- How do we know what was done in studies if they don't all report everything?
- Do we need to adopt comprehensive reporting standards (e.g., https://tms-rat.org :-))

Reviewing / editing / presenting at conferences

- How do we incorporate or refer to best-practice when reviewing others' work?
- Which journals are open to reporting null results, failures to replicate or other critique?
- Which meetings or conferences encourage open dialogue & critique? Which stifle it?
- How can we, respectfully & constructively, criticise dominant paradigms within TMS?

What does a cumulative, theory-driven science of TMS look like?

- When a (famous or popular) TMS method does not work, how do we respond?
- What level of evidence do we require for positive versus negative results?
- When do we end a line of research? When is more research *not* required?
- What will TMS research be like in 2065? What do we *want* it to be like?

Should we align to (eg) Brain Imaging Data Structure (BIDS) standards?

- NIBS-BIDS is coming soon... Is that the best one to follow? (and Big TMS: NIBS-DAS)

TMS@40 Hackathon Briefing: Implementation 1

Assuming that we develop a series of agreements & recommendations about how TMS experiments should be done, how do we make sure that change happens?

This session focuses on what ***TMSMultiLab members*** might do after the workshop...

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Implementation_1

What should you do in your lab?

- If you find that your lab does things differently to another lab, how do you address this?
- Do you feel able to feedback, train, or critique your lab-members (or boss!)?

What should TMSMultiLab do?

- Sharing data, meta-data, code & materials? How, where, what format, at what stage(s) in a study's lifetime?
- Monthly meetings (what topics, who will lead on them)?
- Study design meetings & feedback on project and funding proposals – as part of a monthly meeting, or a separate format?
- Hackathons (how often, what topics, how long & what commitments & resources are needed)?
- Do we need to 'formalise' (e.g., find Officers to run TMSMultiLab for ~1-2 years at a time)

How to acknowledge TMSMultiLab work?

- Can we develop a standard (short!) wording for papers, grants & posters? (why not do it today!?)
- Should TMSMultiLab be a 'group author'? Do we need (e.g.) an email address for it?
- How do we assign authorship in multi-author work (e.g., for the TMS@40 paper)

What role does TMSMultiLab play in university teaching?

- Can we (should we) coordinate across-lab undergraduate & masters projects?
- Can we (should we) coordinate across-year undergraduate & masters projects?
- Can we (should we) develop or share standard teaching resources, slides, etc?

TMS@40 Hackathon Briefing: Implementation 2

Assuming that we develop a series of agreements & recommendations about how TMS experiments should be done, how do we make sure that change happens?

This session focuses on how we can expand ***TMSMultiLab*** & influence researchers who are not already members

EDIT THIS PAGE ONLY:
https://github.com/TMSMultiLab/TMSMultiLab/wiki/TMS@40_Implementation_2

How does TMSMultiLab maintain momentum?

- What 'headlines' do we want to create, based on the TMS@40 workshop?
- Where should we advertise or promote our work?
- Which social media should we use (https://bsky.app/profile/tmsmultilab.bsky.social)
- Which audiences should we try to attract? Which meetings, societies or disciplines?
- Symposia or satellite events at other meetings – which events, when?
- Do we need videos, podcasts, blogs or other multimedia? Who will use them? Who will create them?
- Should we use platforms like PubPeer.com to offer consensus-based critique of others' work? (critique can be positive ;-))

How do we encourage other researchers or students to use the (GitHub) resources?

- Start on the wiki; learn simple html (markdown) coding?
- Promote the wiki as a simple guide to TMS methods?
- Do we need a forum (to deal with queries, e.g., using *Discourse*)? Who will run it?

Should we try to engage Journal Editors with summaries of guidance?

- Which Journals & why? Should we try to get Editorials or Guidance changed?
- Should we produce an 'Editor briefing' – a short digest of the meeting and paper, to send to editors? (the editor of *Cortex* thinks this is a good idea)

Should we link-up with other societies focussed on methods reform?

- Reproducibility networks
- Society for the Improvement of Psychological Science (SIPS)
- Who else?