Document 1 - ICDR

SENEDD STANDARDS COMMITTEE INQUIRY: INDIVIDUAL MEMBER ACCOUNTABILITY (POLITICAL DECEPTION)

RESPONSE TO THE COMMITTEE'S FOLLOW-UP QUESTIONS

Introduction

1. This note responds to the Committee's letter of 2 December 2024. The ICDR is incredibly grateful for the opportunity to offer some answers to the Committee's questions and is delighted to offer its ongoing commitment to assisting the Committee in any way.

2. We have taken account of the evidence provided by others thus far. In view of this, we would like to propose two amendments to the ICDR model. These are not necessarily intended to replace the existing provisions of the ICDR Model but, rather, to provide alternatives and demonstrate how the model can be adapted to take account of different concerns while maintaining it' fundamental integrity:

   (a) **The Gatekeeper Option** – this will make the Standards Commissioner the only person who is empowered to apply for a Correction Order or Disqualification Order. This is intended to take account of concerns about individual citizens being permitted to hold politicians to account under this regime and those who sought to bring the Commissioner into the regime.

   (b) **The Panel Option** – this will provide for the ICDR Model to be administered by an administrative panel rather than the courts.

3. The structure of this note is, therefore, as follows:

   (a) The main body of the note provides answers to the Committee's written questions.

   (b) Appendix 1 sets out the Gatekeeper Option and Panel Option.

   (c) Appendix 2 provides updated suggested legislation for the ICDR Model (taking into account proposed updates and amendments).

   (d) Appendix 3 provides responses to other issues which have arisen over the course of the Committee's inquiry.

4. For convenience, the appendices are attached as a separate document.

<u>Responses to the Committee's Questions</u>

What are your views on how the regulatory model proposed by the IDCR is compliant with the European Convention on Human Rights for example (Article 10) and freedom to stand in an election (Article 3 of Protocol No. 1 - Right to free elections) of the European Convention on Human Rights?

5. The ICDR Model does not impose any new restriction on the Article 10 or A1P3 rights in respect of Members of the Senedd. As set out in the ICDR White Paper, MS are already required to speak truthfully and to correct themselves where they fail to do so. The ICDR Model does not impose a new regulatory burden but, rather, a fairer and more effective mechanism of enforcing an existing duty.

6. It is already well established that the government may place limitations on both the Article 10 and the A3P1 rights in a political context. Political speech is, for example, limited by the National Assembly for Wales (Representation of the People) Order 2007. The law allows for individuals to be disqualified from holding elected office (and for

people to be prevented from standing in an election for a variety of other of other reasons).[1]

7. Article 10 of the ECHR provides that the right to free expression "carries with it duties and responsibilities" and may, therefore be subject:

> *to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society,… for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others…*
>
> .

8. Any measure which impacts on speech is a potential *prima facie* interference with the right to free expression. It is well established, however, that certain limitations on free expression are lawful under the ECHR. Examples of this include defamation laws.[2]

9. The ICDR Model will be permitted by the ECHR because:

   (a) It is "prescribed by law", in that it will be set down in statute;

   (b) It is necessary in a democratic society because,

      (1) Citizens in a democracy are entitled to receive truthful information from politicians.

      (2) For the reasons set out in the ICDR's previous submissions (and by a number of other witnesses), political deception represents a genuine threat to the functioning of democracy and, by extension, to the rights of citizens

[1] For a complete list see: https://www.electoralcommission.org.uk/guidance-candidates-and-agents-local-government-elections-england/what-you-need-know-you-stand-a-candidate/qualifications-and-disqualifications-standing-election/disqualifications (last accessed 9 December 2024).

[2] Defamation Act

which are best realised in such a democracy (as recognised in the preamble of the ECHR).

(c) It is a proportionate measure:

(1) A Correction Order represents only a minor interference with free expression and one that addresses a significant threat to the functioning of democracy (the erosion of public trust in democratic politics caused by political deception) and, by extension, the rights of others.

(2) A Disqualification Order will only arise after the subject has been given every opportunity to correct the record. The court or panel will have a significant degree of discretion about what sanction to impose and will be able to take the requirements of the ECHR into account when deciding what sanction to impose.

10. A3P1 does not confer a stand-alone right to stand in elections. It protects the rights of citizens to:

> …free elections at reasonable intervals by secret ballot, under conditions which will ensure the free expression of the opinion of the people in the choice of legislature.

11. All democratic states disqualify persons from office (and even from voting). This is permitted by A1P3 where they are justified and pursue a legitimate aim.[3] For the reasons set out above, the ICDR Model (and, indeed, the other proposals before the Committee of which the ICDR is aware) meet that test. Indeed, the purpose of these measures is to enhance the freedom of elections by protecting citizens from misleading information and restore public trust in democracy.

---

[3] See, for example, *Py v France* 66289/01, [2005] ECHR 7, [2006] 42 EHRR 26

What are your views on how a new criminal offence would interact with existing offences during an election period, specifically the National Assembly for Wales (Representation of the People) Wales Order 2007, which states that it is illegal to make a false statement concerning the personal character or conduct of a candidate for the purpose of affecting their return as a candidate.

12. Having taken full account of the CBA submissions, the ICDR view is that this measure should not take effect in criminal law. The ICDR Model will work as well, if not better, in Civil Law and Appendix A to this document sets out how it can be given effect as a purely administrative measure. It should, therefore, have no interaction at all.

13. That aside, the 2007 Order offences serve an entirely different purpose from the measure under consideration in this inquiry. The offences in articles 14, 30, and 31 all target, what might be broadly be described as "voter fraud". These are concerned with various forms of casting a ballot fraudulently.

14. Similarly, the offence at section 106 of the Representation of the People Act 1983 also serves a different purpose. It is designed specifically to protect the reputations of candidates.

15. The measure before the Committee is entirely different. It is designed to protect the public from misinformation spread by politicians and to start the process of restoring public trust in politics more broadly. The harm it targets is not focused on the reputations of other politicians but, rather, the representation of basic facts in public discourse. More importantly, the ICDR Model is concerned, first and foremost, with ensuring that misinformation is corrected (thereby empowering citizens to vote based on accurate information and rebuilding public trust in democracy), the question of sanction only arises if a correction is not made.

16. While there may be some (minor) partial overlap between the ICDR Model and the section 106 offence (such as when a false statement is made which impacts on the reputation of another candidate), the two regimes will do entirely different things. The section 106 offence will be concerned with sanction. The ICDR Model measure will be concerned with correction. This overlap is similar to the existing overlap between the

section 106 offence and the law of defamation (which may, in limited circumstances, be relevant to the same incident but concern very different things).

17. The ICDR Model will also act much more quickly, providing a correction in a matter of weeks (and, in many cases, in advance of the relevant election). The section 106 offence will not be tried for many months after the alleged offence has been committed (almost certainly well after the relevant election).

Definitions

Could you provide information on your proposals for defining false statements and why you have chosen this formulation.

18. The language of the ICDR Model was originally based on section 2(2) of the Fraud Act 2006. This was chosen because it is an existing legal definition of a false statement and so one with which the courts will be familiar.

19. On reflection, and having listened to the comments of other respondents, we have now formed the view that a better basis will be section 2 of the Defamation Act 2013. This provides for a court to determine whether a statement is or is not "true" for the purposes of a defamation claim (the truth of a statement is a complete defence to a claim of defamation). We, therefore propose the following language:

> The imputation conveyed by the statement complained of is false or misleading.

20. This draws on the language of both the Fraud Act and the Defamation Act. It is preferable because:

    (a) It acknowledges that the court must first reach a clear decision on the meaning of the Qualifying Statement;

    (b) It reflects the Civil Law in which (it is proposed) the ICDR Model should take effect;

(c) It draws on existing concepts that the courts have a proven record of implementing.

21. A court or administrative panel, seised with a claim for a Correction Order (or other measure, if the ICDR Model is not used), must ask itself:

(a) What is meant by the statement in question? This is a common exercise in defamation proceedings, where the court will determine the meaning of the statement in question in order to decide whether it is defamatory.

(b) Is that meaning "false or misleading"? This requires the court to weigh up the evidence presented by both sides and determine which, given the context of the statement, is more compelling.

What are your views on what evidentiary basis/ material could be required to prove that statements were 'wilfully deceitful' or on the 'balance of probabilities' false.

22. The two phrases quoted mean different things:

(a) A statement is false "on the balance of probabilities" if, having weighed up the evidence presented by both sides, the court considers the evidence for the falsity of the statement outweighs the evidence for the truth of the statement.

(b) A statement is "wilfully deceitful" if:

(1) It is false or misleading (as set out above, assuming the balance of probabilities test is used); and

(2) The maker intended the statement to make people believe something that is false or misleading.

23. A court will determine whether a statement is false or misleading, on the balance of probabilities, in exactly the same way it would determine whether a statement is true on the balance of probabilities in any other civil claim: It will consider the evidence from

both sides, ask itself whether the evidence is convincing and determine which side has the most compelling case.

24. In most cases it is likely that the question will be relatively simple. For example, the question: Is the statement "More illegal immigrants entered the country last year than any year in history" false or misleading would be answered by:

    (a) Identifying the year to which the statement refers (e.g. if it was said in 2024 then "last year" would, to the reasonable person, mean 2023);

    (b) Determining the meaning of "illegal immigrant" (including the legal definition);

    (c) Looking at the publicly available records for each year to determine whether there were more illegal entries in any year other than 2023.

    (d) Considering any other evidence which suggests the publicly available records are inaccurate and determining whether reliance can be placed on them.

25. The ICDR's view is that the "wilfully deceitful" test is inappropriate in this context for the following reasons:

    (a) Other respondents and serving politicians have expressed concerns about a lack of clarity in the phrase "wilfully deceitful". They are concerned they may be sanctioned based on a test they don't understand. The ICDR Model, by focusing on correcting false or misleading statements (and only moving to sanction if correction is refused) avoids this problem.

    (b) If a politician makes a false or misleading statement, then the public will be equally harmed regardless of whether the politician intended to deceive them: the statement is equally false (and, therefore, equally in need of correction) regardless of whether the intent exists. The ICDR Model proposes the non-punitive measure of requiring a simple correction to settle the matter before any sanction is considered.

(c) The phrase "wilfully deceitful" is more appropriate in a criminal law context where the court is considering whether to impose a criminal sanction on a defendant. The ICDR Model is primarily about correction rather than sanction and the ICDR prefers the civil to the criminal model.

26. It would, however, be entirely possible for a court to determine whether a statement is "wilfully deceitful" (although it would, however, be preferable to adopt the existing legal language and describe the statement as "dishonest" - as in section 2 of the Fraud Act 2006 - so that it replicates an existing legal term). When determining cases under the Fraud Act, the court must determine (*inter alia*) whether the defendant intended for the recipient of the statement to be deceived. The courts consider dishonesty in a variety of criminal offences (such as theft). It will conduct the same exercise as in these offences, asking whether the evidence justifies the conclusion (likely on the higher "beyond all reasonable doubt" test) that the defendant was dishonest.

If a new offence on deception were introduced what are your views on what defences should be available to an individual, for example making a false statement for reasons of national security.

27. The ICDR does not think that it is necessary for there to be a criminal offence of deception. A civil regulatory regime would be preferable (and an administrative regime may also work).

28. In any of the above regimes, however, there should be defences available:

29. As to a Correction Notice: There will be times (such as for the purpose of national security, the protection of individuals, or the prevention and detection of crime) that it is necessary for a politician to make a misleading statement. It would, however, likely cause problems down the line if the statute was to attempt an exhaustive list of these (because it is impossible to predict every set of circumstances that will occur in the future).

30. We therefore recommend that the statute adopt a similar exercise to that used in Part II of the Freedom of Information Act 2000 and the Environmental Information Regulations 2004. These require the court to weigh the public interest in favour

of/against making an order. The ICDR Model should adopt a similar test. We suggest a clause which provides for a public interest test and offers a non-exhaustive list of circumstances in which it will apply (this has the added benefit of offering guidance for the court about the level of seriousness of the unlisted circumstances which will qualify). For example:

> (1) The [court/panel] shall not make a Correction Notice if it concludes that in all the circumstances of the case the public interest in not making an order outweighs the public interest in making an order.
>
> (2) Circumstances in which the public interest in not making an order will outweigh the public interest in making an order include but are not limited to:
>
> > (a) It was necessary to make the false or misleading statement for reasons of national security.
> >
> > (b) It was necessary to make the false or misleading statement for the prevention or detection of crime.
> >
> > (c) [Other examples can be added to this list]

31. As to a Disqualification Order: We suggest a defence of "reasonable excuse". This is a common defence to regulatory claims (a number of examples are set out in the ICDR's White Paper). It requires the respondent to provide the court with an explanation for their failure to comply with the Correction Order. This will encompass questions of national security and the prevention and detection of crime but it will also acknowledge that other circumstances may justify a failure to comply. The "reasonable excuse" defence is already built in to the ICDR Model drafting.

Procedure

Could you explain your rationale for allowing registered voters to make complaints directly to the courts and to set out how you propose voters would be informed of any such process, and how it would work in practice.

32. The rationale for this proposal is explained in the ICDR White Paper. For ease of reference, it is repeated here:

> *It is preferable, as a general principle, for politicians to be directly accountable to voters whenever possible. The best way to empower voters in respect of political lying would be to allow any registered voter to apply for a Correction Notice. This is the model used for judicial review and regulatory regimes like the Freedom of Information Act 2000. It will reduce any resource implications and avoid the necessity of adding new tasks to an existing regulatory body. If legislators do not want voters to have direct involvement in the regime then the ICDR Model would also work in the same way if an existing regulatory body (such as the Standards Commissioner or the Public Services Ombudsman).*[4]

33. As set out in the ICDR White Paper and in Appendix 1, the ICDR Model could also function with the sole power to seek a correction notice being given to a regulatory body, such as the Standards Commissioner (as would be the case if the Gatekeeper Option is applied).

34. The way that it would work in practice is set out in the sample clause "Application for A Correction Notice". This was included in the appendices to the White Paper and, for convenience, is also included in Appendix 2 to this paper. In short:

35. Under the "direct access version"

    (a) An applicant must make an application to the County Court/administrative panel using the existing process set out in the Civil Procedure Rules or whatever rules

[4] Fowles, "A Blueprint for Political Honesty", (ICDR, 2024), §39

are adopted by the administrative panel. This involves an application form (which already exists) and a witness statement setting out why the applicant believes a false or misleading statement has been made and exhibiting documentary evidence in support.

(b) If a judge concludes that the application is not vexatious and has a real prospect of success then the papers will be sent to the respondant and they will have the opportunity to put in their own evidence (or may simply agree to correct the record). The court may then make directions for a hearing.

(c) The court will hold a hearing to listen to the evidence and argument and question witnesses (if necessary). The court will determine whether to make a Correction Notice on the basis of the written evidence and evidence given at the hearing.

36. Under the Gatekeeper Option:

(a) A member of the public would make a complaint to the Standards Commissioner.

(b) The Standards Commissioner would be required to evaluate the complaint, decide whether an application for a Correction Notice would (a) have a better than 50% chance of success, and (b) be in the public interest. If the answer to both questions is affirmative then the Commissioner would be required to make an application for a Correction Notice (using the same procedure as above – likely without the necessity of an initial review by a judge).

37. An application for a Disqualification Notice would follow the same procedure. Both sides may put in evidence about whether the respondent (a) complied with the correction notice, and (b) had a reasonable excuse for not doing so.

38. For many areas of law, the government or a relevant organisation publishes informal of formal guidance. It would be sensible to do the same in this case. If the Committee think it appropriate then, it may wish to include a provision requiring the government to provide guidance. A provision of this nature is found at section 182 of the Licensing Act 2003.

What is your response to the view of Transparency International UK that creation of a criminal offence could create risks similar to those of SLAPP (Strategic Litigation against Public Participation) litigation, where those with resources can manipulate the system.

39. The ICDR Model contains stronger safeguards against SLAPP claims than any other legislative scheme (certainly more than the Defamation Act 2013, under which most SLAPP claims are brought). In particular:

    (a) We propose that it will be a criminal offence to make a vexatious claim. A SLAPP would fall squarely within the definition of a vexatious claim.

    (b) An application can be avoided entirely if the politician corrects the record.

    (c) Each application will be reviewed by a judge (or the Standards Commissioner) to determine if it is vexatious before the respondent is even required to respond. The vast majority (arguably all) of SLAPP claims will, therefore, never reach the stage where they require a response.

    (d) The existing costs regime in the civil courts will require the losing party to pay the successful party's costs (the losing party's liability is limited if they concede at an early stage). This will be the case for the civil law version of the ICDR Model and the same rule can be applied to the administrative panel version.

    (e) As a further guard against SLAPP claims, we suggest adding a provision which allows the court to impose costs on the "indemnity basis" (essentially a punitive costs award) if it finds that an application is a SLAPP.

    (f) SLAPP are effective because defamation claims are some of the most expensive litigation (largely because claims take so long so costs run up). The time limits and case management requirements in the ICDR Model mean that this would not happen.

(g) The "Gatekeeper Option" will introduce a new level of protection by only allowing the Standards Commissioner to instigate an application for a Correction Order (see Appendix 1).

In your view, what appeals process should be available as part of any process and what arrangements should be in pace in relation to legal costs, for example where a Member or Candidate contest an allegation of lying successfully.

40. This is set out in the White Paper and the appeals process is already defined in civil law.

41. There should be a right to appeal on a point of law as there is with any other civil or administrative claim. If the regime is given to an administrative panel rather than the courts, then there should be a right to appeal from the decision of the administrative panel on a point of law (see Appendix 1).

42. The ICDR's view on costs is set out above. We also suggest that legal aid (or an equivalent scheme) should be made available.

Sanctions

What are your views on the comments made by the Criminal Bar Association in their written evidence to this Committee, where they stated:

"There are many ways in which available sanctions can be extended in accordance with the Nolan principles, for example, by way of providing for the disqualification of elected Members through the process of recall."

43. The Correction Notice part of the ICDR Model is the most important stage. It would be possible to combine the ICDR Model with recall as a sanction. For example, the penalty for failure to comply with a Correction Notice could be a recall election or petition. This, for the reasons set out below, would not (in the ICDR's view) be as effective as a Disqualification Notice, but it would be better than doing nothing.

44. The chief drawback of recall as a sanction is that it lacks speed. By the time that a recall petition completed, and an election can be organised, the public discourse will have moved on, leaving any correction or consequences meaningless.

45. It also suffers from the drawbacks of the "Pure Politics Model" of regulation, which are set out in the White Paper. For convenience, the relevant section is repeated here:

> ***The Pure Politics Model*** *– This is the model which appears to be preferred by the Senedd Standards Commissioner. It essentially involves "leaving the matter to voters". This model is prima facie attractive but, in reality, does voters a great disservice:*
>
> > ***It's a "Catch-22"*** *- Political lying denies voters the opportunity to make an informed decision. Imposing politicians' false or misleading statements on voters is unfair and undermines confidence in the democratic process.*
> >
> > ***It doesn't work if there's no alternative*** *– Voters distrust all politicians. The unchecked extent of political misrepresentation means that voters do not distinguish between trustworthy and untrustworthy politicians. Telling voters to "just vote for someone else" only works if there's a genuine alternative. At the moment, voters do not see that there is.*
> >
> > ***It doesn't work in practice*** *– Voters only get the opportunity to vote for politicians in the Senedd every five years. It is very difficult to hold politicians accountable at the ballot box for a false or misleading statement if they are not up for election for another half decade. The possibility of recall, which can take up to a year, does not offer substantial mitigation for this problem.*
>
> *The Westminster Parliament, in 2015, introduced recall petitions as a potential sanction for (inter alia) breach of privilege (which can include misleading Parliament). The sanction has not proved effective in respect of political lying:*

Since 2015 trust in politics has declined. The years since 2015 have arguably seen an increase in political dishonesty (and certainly the public perception of political lying).

The process takes too long (there is often more than a year between the false or misleading statement and the recall petition).

Why do you believe that disqualification would be an appropriate sanction if a new criminal offence on deception as introduced.

46. As set out above, the ICDR does not believe a criminal offence is necessary. This question is, therefore, answered from the point of view of a civil or administrative regime (although the same points will apply to a criminal regime).

47. The ICDR believes disqualification is appropriate for the following reasons:

    (a) First, it is what was promised. The Counsel General's statement made specific reference to disqualification. Abandoning that now would send a message to the public that the Senedd is abandoning its commitment and would further damage public trust in the democratic system (and, in particular, the parties currently in the Senedd).

    (b) Second, because, for the reasons set out above, a recall petition is not an appropriate deterrent.

    (c) Third, political deception goes to the heart of a person's suitability for office. We currently disqualify people from holding office if they (for example) have been involved in corrupt practices.[5] A refusal to correct the record after a statement

[5] See https://www.electoralcommission.org.uk/guidance-candidates-and-agents-local-government-elections-england/what-you-need-know-you-stand-a-candidate/qualifications-and-disqualifications-standing-election/disqualifications (last accessed 9 December 2024)

has been proven, in a court of law, to be false or misleading falls squarely within any reasonable person's understanding of corrupt practices.

SAM FOWLES

For the ICDR
16 December 2024

SENEDD STANDARDS COMMITTEE INQUIRY: INDIVIDUAL MEMBER ACCOUNTABILITY (POLITICAL DECEPTION)

RESPONSE TO THE COMMITTEE'S
FOLLOW-UP QUESTIONS

- APPENDICIES -

INDEX TO APPENDICIES

1	Suggested amendments to the ICDR Model
2	Suggested legislative drafting
3	Responses to other matters raised during the inquiry

APPENDIX 1: SUGGESTED AMENDMENTS TO THE ICDR MODEL

Introduction

1. The ICDR welcomes the critiques of the ICDR Model across this Inquiry. While we do not agree with some of them (as set out in the ICDR Response Note – App. 3), we have nevertheless sought to offer some amendments to the ICDR Model which take them into account.

2. The principal concerns appear to be:

   (a) The risk of vexatious claims or SLAPP claims;

   (b) The risk of the courts being overwhelmed;

   (c) The feeling that this matter is inappropriate for the Criminal Courts;

3. The ICDR offers two amendments to address these:

   (a) **The Gatekeeper Option** – The Standards Commissioner will be given the sole power to bring claims for Correction Orders and Disqualification Orders and an explicit power to prosecute those who make vexatious claims. This will provide for a control against trivial and vexatious claims and against the courts being overwhelmed. The Standards Commissioner will be held accountable for their decisions by a right of appeal to the courts.

   (b) **The Panel Option** – The power to make Correction Orders and Disqualification Orders will be given to an administrative panel, overseen by the Standards Commissioner. This will be held accountable by a right to appeal to the courts on a point of law.

The Gatekeeper Option

4. The Standards Commissioner will have the exclusive power to ask the court for a Correction Order or Disqualification Order. The Commissioner will, in effect, act as

the regulator for the rules against political deception and will be empowered to seek court orders as part of that regulatory role. This reflects the powers of other regulators (such as the Information Commissioner) who have exclusive powers to seek certain court orders (including punitive orders) as part of their regulatory role.

5. The ICDR has already expressed our concern that the Standards Commissioner is insufficiently accountable for their decisions and works too slowly. This will be addressed by making the Standards Commissioner accountable to the courts for their decisions. The Standards Commissioner, on receiving a report from a member of the public, will have a duty to seek a Correction Order or Disqualification Order where they conclude:

   (a) It is in the public interest to do so; and

   (b) There is a better than 50% prospect of success.

   This draws on the test applied by the CPS before initiating a prosecution.

6. Given that (as already set out) misleading statements by politicians harm the political discourse and public trust as a whole, it is right that the Standards Commissioner is held accountable for his decisions by a body that the public trusts (the ICDR has already provided the Inquiry with data about the high level of public trust in the courts). Many public bodies are accountable to the courts for their regulatory decisions. Alcohol licensing authorities, for examples, can have many of their decisions reviewed by the courts. We therefore propose that, should the Standards Commissioner decline to seek a relevant order, having received a complaint from a member of the public, that person should have the right to ask the courts to review that decision. There is no need for a similar review provision if the Commissioner does seek a relevant order because, under the law as it stands, the respondent will already be able to ask the court to strike out the application as an abuse of process or seek a judicial review.

7. For the reasons set out in the ICDR's previous submissions, it is vital that these proceedings happen quickly. For that reason, we recommend imposing statutory time

limits on the Standards Commissioner's decision-making in this area (it is relatively common for regulators to be subject to time limits).

8. There will be some resource implications for this system. It adds to the work of the Standards Commissioner and will require relatively swift decision-making. It will likely be necessary for the Standards Commissioner to appoint a number of caseworkers to handle the workload and to instruct lawyers to handle applications to the court. For the reasons already set out by the ICDR, the number of complaints is likely to be relatively low, so the resource implications are likely to be commensurately relatively low.

9. A sample clause giving effect to the above can be found at Appendix 1 of this Note.

<u>The Administrative Panel Option</u>

10. Should the Committee decide that it would prefer to keep the matter out of the courts so far as possible, we have also given some thought to how the ICDR Model could work if enforced by an administrative panel.

11. While the comments given in the ICDR White Paper, the Rebuttal Note, and by Sam Fowles in oral evidence, which address the advantages of the court system (in terms of independence, transparency, speed, and public trust) we nevertheless wanted to give the Committee an option that would work in an administrative context.

12. We therefore propose the following:

    (a) The Senedd creates (by statute) a "Standards Panel". If the Standards Commissioner does not have a role in bringing claims for Correction Orders or Disqualification Orders, then this may be set up under the auspices of the Commissioner. If the Commissioner does have such a role, then it should be independent.

(b) The Standards Panel will have the power to make Correction Orders and Disqualification orders on the same terms as proposed for the courts in the White Paper.

(c) The Standards Panel would sit with a legally qualified chair and two lay "wing members". This replicates the constitution of the First Tier Tribunal and many disciplinary panels (such as police disciplinary panels). It would be sensible to recruit a pool of legally qualified chairs and lay members to ensure availability. Panel members would likely work on a "fee paid" basis (as with members of police disciplinary panels) rather than a full-time basis so would only need to be paid for the time they work.

(d) The Panel should be required by statute to adopt procedure rules which offer an equivalent standard of procedural fairness to the rules of the most appropriate court (it is suggested that the First Tier Tribunal would be an appropriate comparator). Alternatively, such rules could be set down by secondary legislation.

(e) To ensure the panel is accountable for its application of the rules, there should be a right of appeal, on a point of law, to an appropriate court (it is suggested either the County Court or the High Court would be appropriate).

SAM FOWLES
For the ICDR
6 December 2024

APPENDIX 2: SAMPLE LEGISLATION

Please not that these clauses are intended of examples of how sample legislation could look. They are intended to offer a starting point for discussion, not a finished product.

The Gatekeeper Option

Standards Commissioner's Duty

(1) On receiving a complaint from a member of the public or otherwise becoming aware that:

(a) A Qualifying Person has made a Qualifying Statement; and

(b) The Qualifying Statement is false or misleading;

The Standards Commissioner shall consider whether to make an application for a Correction Notice.

(2) On receiving a complaint from a member of the public or otherwise becoming aware that a person who is subject to a Correction Notice has failed to comply with the requirements of that notice within the time specified in the notice the Standards Commissioner shall consider whether to make an application for a Disqualification Notice.

(3) The Standards Commissioner shall make an application for a Correction Notice, or a Disqualification notice if he forms the view that:

(a) There is a public interest in making the application; and

(b) The application has a better than fifty percent prospect of success.

(4) Within seven days of receiving a complaint under this section the Standards Commissioner must:

(a) Apply for an order under [relevant subsequent provisions – Correction Order/Disqualification Order]; or

(b) Publish a decision notice explaining why he has decided not to seek an order under [relevant subsequent provisions – Correction Order/Disqualification Order]; or

(c) Publish a decision notice explaining that exceptional circumstances mean that it is not possible for the Standards Commissioner to do either (a) or (b) in the time available and stating the reasons for this and the date by which he will do either (a) or (b).

Appeal against a decision of the Standards Commissioner

(1) A person who:

(a) Has made a complaint to the Standards Commissioner; and

(b) Is dissatisfied with the Standards Commissioner's response to that complaint

may appeal to the [insert relevant court].

(2) An appeal under subsection (1) may not be made unless more than seven days have passed since the Standards Commissioner received the complaint.

(3) On receipt of an appeal under this section the [insert relevant court] may exercise any power that the Standards Commissioner could have exercised on receipt of the original complaint.

Power to issue a Correction Notice

(1) This provision applies when a Qualifying Person makes a Qualifying Statement.

(2) A Qualifying Person is a person who is a Member of the Senedd or a candidate for election to the Senedd.

(3) A Qualifying Statement is a statement that:

(a) Is published; and

(b) A reasonable person could understand to be a statement of fact;

(c) A statement of fact shall not include a statement of honest opinion.

(4) Where [a court/the Standards Panel] finds that a Qualifying Person:

(a) Has made a Qualifying Statement; and

(b) The statement was made in the course of their conduct as a Qualifying Person or there is a public interest in whether the statement is false or misleading; and

(c) The imputation conveyed by the statement complained of is false or misleading;

It shall make and issue a Correction Notice to the in respect of that statement.

(5) A Correction Notice is a notice which requires the Qualifying Person named therein to correct the Qualifying Statement specified in the notice in the manner specified in the notice.

(6) [the court/the Standards Panel] shall not issue a correction notice if:

(a) The false or misleading aspect of the Qualifying Statement is trivial; or

(b) It was necessary to make the Qualifying Statement including the false or misleading aspect for the purposes of national security or law enforcement.

(7) A person named in a Correction Notice must comply with the notice by publishing the required correction in the same forum as they made the statement specified in the Notice within seven days of the date of the Notice.

(8) If it is not reasonably possible to make the required correction in the same forum as the Qualifying Statement was made then they must make the correction in a forum of equivalent or greater prominence.

(9) The Standards Commissioner shall cause the Correction Notice to be published on a public register which is available online and to be published in hard copy and online in a newspaper which is of national prominence in Wales.

(10) Appeal against a decision to issue a Correction Notice lies to the Crown Court and may only be made on a point of law.

(11) The standard of proof for any matter in this section shall be the balance of probabilities. Where [a court/the Standards Panel] finds that there is a real prospect that:

(a) A Qualifying Statement is false or misleading; or

(b) Did not have an evidential basis at the time that it was made

The evidential burden of proving the truth of the Qualifying Statement shall lie with the person who made that statement.

[ALTERNATIVELY: (11) The standard of proof for any matter in this section shall be the balance of probabilities]

(12) [a court/the Standards Panel] exercising any power under this section must acknowledge its duty under section 3 of the Human Rights Act 1998.

(13) The [court/panel] shall not make a Correction Notice if it concludes that in all the circumstances of the case the public interest in not making an order outweighs the public interest in making an order.

(14) Circumstances in which the public interest in not making an order will outweigh the public interest in making an order include but are not limited to:

(a) It was necessary to make the false or misleading statement for reasons of national security.

(b) It was necessary to make the false or misleading statement for the prevention or detection of crime.

(c) [Other examples can be added to this list]

(14) Definitions

a. A candidate for election to the Senedd has the same meaning as in section 7 of the Government of Wales Act 2006.
b. A statement is "published" if it is a "publication" according to the law of defamation.
c. "Honest opinion" shall have the same meaning as in the Defamation Act 2013.

Application for a Correction Notice

(1) An application for a Correction Notice may be made by any person who is registered to vote in Wales.

ALTERNATIVELY: An application for a Correction Notice may be made by [INSERT APPROVED AUTHORITIES]

(2) An application under this part must be made by the same process as is used for starting a prosecution.

(3) An application under this part must be served on the respondent within 24 hours of the information being laid.

(4) An application under this part must be heard by a magistrates' court no later than seven days after it is served on the respondent.

(5) A District Judge sitting in the magistrates' court must review an application under this part on the papers within [x] hours of the information being laid.

(6) An application under this part must be dismissed without a hearing if a District Judge determines that it is vexatious or has no real prospect of success.

(7) An application may not be made under this section after:

(a) More than six months have passed since the date on which the relevant Qualifying Statement was made; or

(b) More than six months have passed since the date on which sufficient evidence to establish on the balance of probabilities that the relevant Qualifying Statement was false or misleading or contained false or misleading aspects became available to the public.

Whichever is longer.

(8) Where it is reasonably necessary for a party to an application under this part to rely on information to which one of the exemptions in Part II of the Freedom of Information Act would otherwise apply:

(a) [the court/the Standards Panel] may exclude any person before considering that information.

(b) Where [the court/the Standards Panel] excludes a party to the application that party may be represented by a special advocate during the parts of the hearing from which they are excluded.

Failure to comply with a correction notice

(1) If [a court/the Standards Panel] is satisfied beyond all reasonable doubt that person named in a Correction Notice failed without reasonable excuse to comply with that notice within the time specified then any person who is registered to vote in Wales may apply for a Disqualification Order in respect of the person named in the notice.

(2) If on an application under sub-section (1) [a court/the Standards Panel] finds that a person named in a Correction Notice has not complied with the notice within a specified time then [the court/the Standards Panel] must make a Disqualification Order.

(3) An application under subsection (1) must be accompanied by a witness statement.

(4) A Disqualification Order shall prohibit the subject from being a member of the Senedd until such a time as shall be specified in the order.

(5) The amount of time specified in the order must be at least as long as the maximum period of time before the next Senedd election.

(6) In determining the length of the Disqualification Order [the court/the Standards Panel] shall have regard to:

(a) Any harm caused by the Qualifying Statement;

(b) The extent to which the Qualifying Statement was published;

(c) Whether the defendant should reasonably have been aware that the Qualifying Statement was false or misleading or included false or misleading elements;

(d) Whether the defendant has previously been convicted of an offence under this section or has been the subject of a Correction Order.

(7) Schedule 1A of the Government of Wales Act 2006 is amended to add the following words:

"(9) A person who is subject to a Disqualification Order which has not expired."

Offence of making a vexatious application [ALTERNATIVELY: report]

(1) It shall be an offence to make a vexatious application [ALTERNATIVELY: report with the intention that it shall cause or influence an appropriate authority to make an application under this part.

(2) The offence in (1) shall be triable summarily and shall be punishable by a fine.

Power to take over applications

(1) [Insert relevant public authority] may take over an application for a Correction Order or an application for a Disqualification Order at any stage.

Administrative Panel

Standards Commissioner's Duty

(1) On receiving a complaint from a member of the public or otherwise becoming aware that:

(a) A Qualifying Person has made a Qualifying Statement; and

(b) The Qualifying Statement is false or misleading;

The Standards Commissioner shall consider whether to make an application for a Correction Notice.

(2) On receiving a complaint from a member of the public or otherwise becoming aware that a person who is subject to a Correction Notice has failed to comply with the requirements of that notice within the time specified in the notice the Standards Commissioner shall consider whether to make an application for a Disqualification Notice.

(3) The Standards Commissioner shall make an application for a Correction Notice, or a Disqualification notice if he forms the view that:

(a) There is a public interest in making the application; and

(b) The application has a better than fifty percent prospect of success.

(4) Within seven days of receiving a complaint under this section the Standards Commissioner must:

(a) Apply for an order under [relevant subsequent provisions – Correction Order/Disqualification Order]; or

(b) Publish a decision notice explaining why he has decided not to seek an order under [relevant subsequent provisions – Correction Order/Disqualification Order]; or

(c) Publish a decision notice explaining that exceptional circumstances mean that it is not possible for the Standards Commissioner to do either (a) or (b) in the time available and stating the reasons for this and the date by which he will do either (a) or (b).

Appeal against a decision of the Standards Commissioner

(1) A person who:

(a) Has made a complaint to the Standards Commissioner under this part; and

(b) Is dissatisfied with the Standards Commissioner's response to that complaint

may appeal to the [insert relevant court].

(2) An appeal under subsection (1) may not be made unless more than seven days have passed since the Standards Commissioner received the complaint.

(3) On receipt of an appeal under this section the [insert relevant court] may exercise any power that the Standards Commissioner could have exercised on receipt of the original complaint.

The Standards Panel

(1) There shall be a body called the Standards Panel.

(2) The Standards Panel shall have the powers set out in sections [insert as appropriate].

(3) The Standards Panel must conduct itself according to procedural rules which provide for a standard of procedural fairness which is equivalent to that conferred by the [for example Civil Procedure Rules].

(4) The Standards Panel shall sit with three members. These shall be:

(a) A person who is a member of the Bar of England and Wales or who is a qualified solicitor.

(b) Two persons who are not legally qualified.

(5) All members of the panel must have appropriate professional experience and be able to demonstrate a high level of integrity.

(6) The legally qualified member of the panel shall act as the chair.

Power to issue a Correction Notice

(1) This provision applies when a Qualifying Person makes a Qualifying Statement.

(2) A Qualifying Person is a person who is a Member of the Senedd or a candidate for election to the Senedd.

(3) A Qualifying Statement is a statement that:

(a) Is published; and

(b) A reasonable person would understand to be a statement of fact.

(c) For the purposes of this section a statement of fact shall not include a statement of honest opinion.

(4) Where the Standards Panel finds that a Qualifying Person:

(a) Has made a Qualifying Statement; and

(b) The statement was made in the course of their conduct as a Qualifying Person or there is a public interest in whether the statement is false or misleading; and

(c) The Qualifying Statement in question is not true or is misleading;

It shall make and issue a Correction Notice to the in respect of that statement.

(5) A Correction Notice is a notice which requires the Qualifying Person named therein to correct the Qualifying Statement specified in the notice in the manner specified in the notice.

(6) The Standards Panel shall not issue a correction notice if:

(a) The false or misleading aspect of the Qualifying Statement is trivial; or

(b) It was necessary to make the Qualifying Statement including the false or misleading aspect for the purposes of national security or law enforcement.

(7) A person named in a Correction Notice must comply with the notice by publishing the required correction in the same forum as they made the statement specified in the Notice within seven days of the date of the Notice.

(8) If it is not reasonably possible to make the required correction in the same forum as the Qualifying Statement was made then they must make the correction in a forum of equivalent or greater prominence.

(9) The Standards Commissioner shall cause the Correction Notice to be published on a public register which is available online and to be published in hard copy and online in a newspaper which is of national prominence in Wales.

(10) Appeal against a decision to issue a Correction Notice lies to the Crown Court and may only be made on a point of law.

(11) The standard of proof for any matter in this section shall be the balance of probabilities. Where the Standards panel finds that there is a real prospect that:

(a) A Qualifying Statement is false or misleading; or

(b) Did not have an evidential basis at the time that it was made

The evidential burden of proving the truth of the Qualifying Statement shall lie with the person who made that statement.

[ALTERNATIVELY: The standard of proof for any matter in this section shall be the balance of probabilities]

(12) The Standards Panel exercising any power under this section must acknowledge its duty under section 3 of the Human Rights Act 1998.

(14) Circumstances in which the public interest in not making an order will outweigh the public interest in making an order include but are not limited to:

(a) It was necessary to make the false or misleading statement for reasons of national security.

(b) It was necessary to make the false or misleading statement for the prevention or detection of crime.

(c) [Other examples can be added to this list]

(14) Definitions

a. A candidate for election to the Senedd has the same meaning as in section 7 of the Government of Wales Act 2006.
b. A statement is "published" if it is a "publication" according to the law of defamation.
c. "Honest opinion" has the same meaning as in the Defamation Act 2013.

Application for a Correction Notice

(1) An application for a Correction Notice may be made by any person who is registered to vote in Wales.

ALTERNATIVELY: An application for a Correction Notice may be made by [INSERT APPROVED AUTHORITIES]

(2) An application under this part must be made by the same process as is used for starting a prosecution.

(3) An application under this part must be served on the respondent within 24 hours of the information being laid.

(4) An application under this part must be heard by a magistrates' court no later than seven days after it is served on the respondent.

(5) A District Judge sitting in the magistrates' court must review an application under this part on the papers within [x] hours of the information being laid.

(6) An application under this part must be dismissed without a hearing if a District Judge determines that it is vexatious or has no real prospect of success.

(7) An application may not be made under this section after:

(a) More than six months have passed since the date on which the relevant Qualifying Statement was made; or

(b) More than six months have passed since the date on which sufficient evidence to establish on the balance of probabilities that the relevant Qualifying Statement was false or misleading or contained false or misleading aspects became available to the public.

Whichever is longer.

(8) Where it is reasonably necessary for a party to an application under this part to rely on information to which one of the exemptions in Part II of the Freedom of Information Act would otherwise apply:

(a) The Standards Panel may exclude any person before considering that information.

(b) Where the Standards Panel excludes a party to the application that party may be represented by a special advocate during the parts of the hearing from which they are excluded.

Failure to comply with a correction notice

(1) If the Standards Panel is satisfied beyond all reasonable doubt that person named in a Correction Notice failed without reasonable excuse to comply with that notice within the time specified then any person who is registered to vote in Wales may apply for a Disqualification Order in respect of the person named in the notice.

(2) If on an application under sub-section (1) the Standards Panel finds that a person named in a Correction Notice has not complied with the notice within a specified time then the Standards Panel must make a Disqualification Order.

(3) An application under subsection (1) must accompanied by a witness statement.

(4) A Disqualification Order shall prohibit the subject from being a member of the Senedd until such a time as shall be specified in the order.

(5) The amount of time specified in the order must be at least as long as the maximum period of time before the next Senedd election.

(6) In determining the length of the Disqualification Order the Standards Panel shall have regard to:

(a) Any harm caused by the Qualifying Statement;

(b) The extent to which the Qualifying Statement was published;

(c) Whether the defendant should reasonably have been aware that the Qualifying Statement was false or misleading or included false or misleading elements;

(d) Whether the defendant has previously been convicted of an offence under this section or has been the subject of a Correction Order.

(7) Schedule 1A of the Government of Wales Act 2006 is amended to add the following words:

"(9) A person who is subject to a Disqualification Order which has not expired."

Offence of making a vexatious application [ALTERNATIVELY: report]

(1) It shall be an offence to make a vexatious application [ALTERNATIVELY: report with the intention that it shall cause or influence an appropriate authority to make an application under this part.

(2) The offence in (1) shall be triable summarily and shall be punishable by a fine.

APPENDIX 3: RESPONSE TO OTHER MATTERS ARISING DURING THE INQUIRY

Should Correction Notices have a higher standard of proof?

13. *B (Children)*[1] the (then) House of Lords rejected the principle that more serious allegations require a higher standard of proof when a protective order is concerned. As Lady Hale (with whom the other members of the curt agreed) put it:

> *Lord Lloyd, at pp 577–578 on the other hand, took a more straightforward line: "In my view the standard of proof under [section 31(2)] ought to be the simple balance of probability however serious the allegations involved … mainly because section 31(2) provides only the threshold criteria for making a care order …*
>
> *My Lords, Lord Lloyd's prediction proved only too correct. Lord Nicholls's nuanced explanation left room for the nostrum, "the more serious the allegation, the more cogent the evidence needed to prove it", to take hold and be repeated time and time again in fact-finding hearings in care proceedings: see, for example, the argument of counsel for the local authority in In re U (A Child) (Department for Education and Skills intervening) [2005] Fam 134 , 137. It is time for us to loosen its grip and give it its quietus.*[2] [Emphasis added]

14. The *ratio* of the court's decision was that a care order is intended to protect the child, not to punish anyone (although the parents in question may well see it as a punishment). The Correction Order is of entirely the same nature. It is intended to protect the public from false information. It is not intended to punish the politician who makes the correction (although, as with the parents of a child taken into care, they may subjectively view it as such). The aspect of the ICDR Model which is intended to punish and deter is the Disqualification Order. The ICDR agrees that if a criminal law route is considered appropriate (although the ICDR believes the civil route is preferable) then the standard of proof for a Disqualification Order should be "beyond reasonable doubt".

Does the ICDR Model sanction "deliberate deception" or just "deception"?

[1] [2008] UKHL 35, [2008] 4 All ER 1, [2009] 1 AC 11

[2] Re B (Children) (Care Proceedings: Standard of Proof) [2009] 1 A.C. 11

15. The ICDR Model is clear that only deliberate deception will be sanctioned. As the White Paper sets out:

> *If a MS or candidate fails to comply with a correction notice within seven days, then they will have demonstrated that:*
>
> > *(a) They are aware that, on the balance of probabilities and given the best evidence, their statement was false [*or misleading*];*
> >
> > *(b) They have been given an opportunity to correct the statement and have not done so.*
> >
> > *(c) They knew they were liable to sanction if they failed to correct the statement and have nevertheless chosen not to correct it.*
>
> *It is, therefore, fair to impose the sanction at this point. It is, therefore, right that the court should have the power to impose sanction (see below) at this point.*

Does the ICDR Model allow claims to be brought "without any evidence"?[3]

16. All applications for Correction Orders (and Disqualification Orders) must be supported by evidence. Under the ICDR Model a person wishing to bring a claim must:

    (a) Establish that it is not vexatious and has a real prospect of success;

    (b) Establish that the defendant is a "Qualifying Person";

    (c) Establish that the statement in question is a "Qualifying Statement";

    (d) Establish, on the balance of probabilities (i.e. by adducing evidence), that there is a "real possibility" that the statement is false or misleading or did not have an evidential basis at the time that it was made.

17. Only if all the above can be proven, will the politician be required to prove, on the balance of probabilities, the truth of their statement.

18. This is well established in regulatory matters. Examples include:

[3] Meeting 18 November 2024, Transcript, 220

(a) Companies Act 2006, s. 451;

(b) Health and Safety at Work Act, ss. 2, 3, and 40;

(c) Criminal Justice Act 1988, s. 139

(d) Female Genital Mutilation Act 2003, s. 3A

19. It is well established that regulatory offences can place the burden on the defendant to prove that they had a reasonable excuse for failing to comply with a regulatory requirement. For example:

(a) Town and Country Planning Act, s. 179;

(b) Housing Act 2004, ss. 30, 32, and 72.

Would the ICDR Model require a change to the Criminal Procedure Rules?[4]

20. The ICDR Model will not take effect in criminal law. It will not require a change to the civil procedure rules (as set out above).

21. If Magistrates Courts were to be given responsibility for enforcing the regime, then the existing procedure rules would remain in place and would not need to be changed. Magistrates Courts use the Criminal Procedure Rules ("**CrimPR**") for both criminal and non-criminal cases.

22. Appendix 2 of the ICDR White Paper makes clear that the procedure for seeking a Correction Order (if the Magistrates Court Route is preferred) should be conducted in accordance with the existing Criminal Procedure Rules;

> *(2) An application under this part must be made by information to a magistrates' court in Wales and must be accompanied by a witness statement.*

23. Part 7 of the Criminal Procedure rules provides:

[4] "Briefing Note" §§28-31

Note. In some legislation, including the Magistrates' Courts Act 1980, an application for the issue of a summons or warrant is described as an 'information' and serving an application on the court officer or presenting it to the court is described as 'laying' that information.

24. If that is not sufficiently clear from the drafting of the sample clause then it can be clarified (and has been in Version 2 of the White Paper). It may also be helpful to specify that the requirements of Part 8 of the CrimPR (initial details of the prosecution case) should also be complied with. This would put a substantial burden on the person seeking a Correction Order. Such a burden seems to be appropriate in this context.

Is the single judge review stage a new invention?

25. Before any prosecution can be brought, the prosecutor must apply to the Magistrates' Court for a summons. A relevant judge (either a magistrate or a District Judge) must review the application and determine whether it meets the test to begin a prosecution. The High Court has confirmed that the threshold for issuing a summons is, for offences of this kind, "a high one".[5]

26. It is, therefore, established that a judge will review any criminal prosecution to ensure that it meets basic thresholds. In respect of certain matters, this must happen quickly (see, for example, Anti-social Behaviour Crime and Policing Act 2014, s. 80(3)). In judicial review, in the High Court, a High Court judge will conduct a process which is almost identical to that proposed in the ICDR Model before permission is given to bring a judicial review. This can be done on an urgent basis (including within 24 hours).

[5] See, for a summary of the position, the Westlaw summary of ***R (Johnson) v Westminster Magistrates* [2019] EWHC 1709 (Admin),** available at https://uk.westlaw.com/Document/ID8F008309D9811E993F0E72A9C5818AD/View/FullText.html?navigationPath=Search%2Fv1%2Fresults%2Fnavigation%2Fi0a89c37e0000019340c5d31a0399ca43%3Fppcid%3D54514ed1fc56477282e36200bbdb9dc7%26Nav%3DUK-CASES%26fragmentIdentifier%3DID8F008309D9811E993F0E72A9C5818AD%26parentRank%3D0%26startIndex%3D1%26contextData%3D%2528sc.Search%2529%26transitionType%3DSearchItem&listSource=Search&listPageSource=169837ea4c6ce983d8dbba77ef85301b&list=UK-CASES&rank=1&sessionScopeId=1e51a13c88b52332ce2678ffe7dd79ae8eec510ee3c4e6176f3c34a2ecb431fb&ppcid=54514ed1fc56477282e36200bbdb9dc7&originationContext=Search%20Result&transitionType=SearchItem&contextData=(sc.Search)&comp=wluk&navId=FEC1646E31EA8A31E6B54EEB0AD4BD73 (last accessed 18 November 2024)

27. The ICDR Model provides for a hybrid of these. It doesn't replicate either exactly but draws on concepts familiar to lawyers and judges.

Does the ICDR Model lack the safeguards in place in other areas of the law?

28. This criticism appears to be targeted at the version of the ICDR Model whereby any registered voter in Wales would be entitled to bring a claim for a Correction Notice. It should be noted that the Model proposes, in the alternative, that the power be limited to the CPS or an alternative regulatory body. This would appear to be a complete answer to such criticism.

29. It is, however, simply wrong to say that the "registered voter" version does not contain the procedural safeguards present in other areas of criminal law:

    (a) Any citizen can bring a private prosecution in England and Wales (save for in very limited circumstances). The Attorney General has a power to "take over" (and, if it so chooses, discontinue) private prosecutions. Such a power could be extended to applications for Correction Notices and subsequent disqualification claims. Version 2 of the ICDR Model incorporates this amendment.

    (b) The ICDR Model addresses the fact that any citizen can bring a claim by imposing stricter measures to block vexatious claims at the application stage. These include:

        (1) The power for a judge to dismiss claims which are vexatious or have no prospect of success on the papers;

        (2) Making a vexatious application will be a criminal offence in and of itself;

        (3) Where a claim is not vexatious, the claimant will still be required to prove their case. A court, under existing rules, will have a discretion to impose a cost sanction if the claim fails.

Does the common law offence of "misconduct in public office" provide a satisfactory alternative?

30. First, the Law Commission has investigated the efficacy of the common law offence and found that "due to the concerns with the terms and practical effects of the common law offence", it should be abolished in its current form. The Law Commission has recommended that it be replaced with two statutory offences which deal with (a) corruption and (b) breach of duty that leads to death or serious injury. These are clearly not appropriate to deal with the instant matter.[6]

31. Second, the common law offence has already been tested in the context of political deception and the Divisional Court made clear that it is not an appropriate instrument. In ***R (Johnson) v Westminster Magistrates*** **[2019] EWHC 1709 (Admin)**, the High Court considered whether Boris Johnson could lawfully be prosecuted under the common law offence for allegedly false statement(s) made in the context of the Brexit campaign. Lord Justice Rafferty and Mr Justice Supperstone held (*inter alia*):

> *Mr Coppel accepts that there is no precedent for any office holder being prosecuted for misconduct in public office for wilfully making/endorsing a misleading statement in and for the purposes of political campaigning, or even any comparable case. He is thus obliged to submit that this matters not since what is alleged falls within the principles applicable to the offence. It does not.*
>
> And
>
> *All the cases to which we have referred and many more we were shown share <u>the common feature of corrupt abuse of public power for personal gain</u>, or gross neglect in failing to comply with the core duties of the office. Such conduct is capable of satisfying the connected tests of breach of duty and the gravity necessary for the offence to be established. The offence will be made out only if the manner in which the specific powers or duties of the office are discharged brings the misconduct within its ambit. Consequently, at the time of the alleged misconduct the individual must be acting as, not simply whilst, a public official.* [Emphasis added]

32. Third, the common law offence is targeted at corrupt abuse of an office. It does not apply to candidates or to members of parliament when they are campaigning. Political deception is an entirely different beast. Any mechanism to combat political deception must deal with statements made during campaigns and must encompass candidates.

[6] https://lawcom.gov.uk/project/misconduct-in-public-office/ (last accessed 24 November 2024)

33. Fourth, the common law offence targets an entirely different harm ("corrupt abuse of public office for personal gain"). The instant issue is not about misuse of office or personal gain. It is to protect the public against false or misleading statements made by politicians. The common law offence should not be twisted into something for which the courts have made clear it is ill suited.

34. Fifth, the common law offence applies in England and Wales. Attempting to both amend the offence and set it on statutory footing such that it only applies in Wales would mean that the same offence has different meanings on either side of the border. This is a recipe for confusion.

35. Sixth, the common law offence suffers from the very problem that the ICDR Model is designed to avoid. There is no clarity as to what it means (in the common law offence) for a politician to "wilfully" make a false or misleading statement. There is no justification for adopting a measure with this lack of clarity.

Is the ICDR Model compatible with the European Convention on Human Rights?

36. First, the requirement for the politician to prove the truth of their statement does not arise in respect of a criminal sanction. It only arises at the Correction Notice stage. The ICDR does not propose that a Correction Notice is treated as a criminal sanction (any more than a planning enforcement notice, closure notice, of information notice is a criminal sanction). It, rather, provides for a regulatory step which allows the subject to correct the record while avoiding criminal sanction. The Magistrates Courts regularly administer regulatory procedures that are not criminal sanctions (see, for example, proceedings under the Licensing Act 2003).

37. It is not clear why politicians would object to saying "I made an error, please allow me to correct the record". Indeed, doing so would likely increase public trust in politicians (polling suggests only 9% trust politicians to tell the truth – the lowest level in recorded history).[7]

[7] https://www.ipsos.com/en-uk/ipsos-trust-in-professions-veracity-index-2023 (last accessed 19 November 2024)

38. Second, a politician would only be required to prove the truth of their statement if the claimant first established that there is a "real prospect" (formerly "real possibility" – see below) that the statement is false or misleading. It is well established that the burden of proving that statement falls on the maker. For example, section 2 of the Defamation Act 2013 requires a that a person relying on the defence of "truth" must show that "the imputation conveyed by the statement complained of is substantially true."

39. Third, even if the ICDR Model did impose a reverse burden in respect of a criminal sanction, the criticism would still be wrong. Both the European Court of Human Rights and the (then) House of Lords (now Supreme Court) have already ruled on this point:

40. In ***Salabiaku v France* (1988) 13 EHRR 379**, the ECtHR held (in respect of reverse burdens):

> *Clearly the Convention does not prohibit such presumptions in principle.*

41. In ***Jobe v UK* (2011) 53 E.H.R.R. 17** the ECtHR held that a provision which placed the burden on the defendant to prove that he had a "reasonable excuse" did not violate the Convention.

42. The House of Lords/Supreme Court has also held that a reverse burden can be interpreted compatibly with the Convention.[8] This matter would not arise in judicial consideration of the ICDR Model but, if it did, then the courts could apply it in a manner that is compatible. Version 2 of the ICDR Paper contains a small amendment to make this clear.

Will the ICDR Model lead to a large number of claims?[9]

43. The ICDR is accused of not conducting any research on the likely volume of claims. The ICDR relies on the evidence already before the Standards Committee which demonstrates that the "floodgates" argument has never been borne out.

[8] For example: ***R (Kebilene) v Director of Public Prosecutions* [2000] 2 AC 326**; ***R v Lambert* [2001] UKHL 37**; ***R v Carass* [2001] EWCA Crim 2845**

[9] "Briefing Note", §32

44. It may assist the Committee to be reminded that it is already in receipt of analysis, from the Senedd's own lawyers, which details similar provisions in other jurisdictions, none of these have resulted in a "floodgates" situation.

45. In previous debates the comparison was drawn with the Hate Crime and Public Order (Scotland) Act 2021. Despite similar predictions, reports under that Act never reached overwhelming numbers. There are around 22 reports per week under the 2021 Act[10] in a country of 5.3 million people (i.e. 5.4 million potential defendants). For the vast majority of the time, there will only be 60-90 potential respondents to a Correction Notice claim (i.e. the number of MS). As an illustration, if the rate of complaint for the 2021 Act were to be repeated in respect of the ICDR Model, we can expect a rate of 0.00066 claims per week before 2026.

<u>Does the ICDR Model risk "politicising the courts"?</u>

46. The courts are to be preferred as a tribunal precisely because they are non-political. Polling by the Constitution Unit at University College London shows that most people believe that the courts should play an important role in ensuring that politicians "operate within the rules".

47. Judges rank fourth in the league table of professions who people in the UK trust to tell the truth (70% trust judges "a great deal" or "a fair amount"). Politicians (Members of Parliament) rank second last (76% trust them "not much" or "not at all"). This is despite a sustained period of argument, by governments found to have broken the law, that judges who rule against them are "political".

48. The British public is clearly intelligent enough to distinguish between "interfering in politics" and "enforcing the rules". The suggestion that allowing courts to arbitrate on matters of fact and law will somehow "politicise" them, mistakes a political attack line for substantial analysis.

[10] Not including anonymous reports, which would not be possible under the ICDR Model

Is there a risk of "lawfare"?

49. It has been suggested that the ICDR Model would somehow lead to US-style electoral lawfare. The comparison between the UK and US systems on this point is, with respect, a poor one. In the US, judges are politically appointed. This makes the courts a ground of political combat. That is not the case in the UK. Judges in the UK are not only politically neutral as a matter of fact and law but, as set out above, are clearly trusted as such by the public.

What are the drawbacks of an administrative procedure?

50. Various "administrative" alternatives have been proposed which involve either expanding the Standards Commissioner's powers or setting up some form of internal tribunal. None of these have been fleshed out so it is difficult to comment on them in detail. There are, however, some clear issues of principle:

    (a) **They lack public trust** – The evidence above clearly demonstrates that the public trust the courts. Giving the courts the role of enforcing the rules on truth-telling means that the public can have confidence that it is in the hands of a trusted institution. Giving it to an administrative body will likely be seen by the public as empowering another quango.

    (b) **They lack independence** – All of the proposals either involve a procedure that is ultimately subject to political control (like the existing Standards Commissioner system) or sits within the existing institution. These will either lack independence or be seen as lacking independence.

    (c) **They lack accountability** – None of the administrative proposals set out how those sitting in judgment will be accountable for their decisions. Independent courts are accountable because judges' decisions can be overturned by more senior courts.

    (d) **They lack clarity and predictability** – None of the administrative proposals offer the clarity that comes with the doctrine of precedent applied in courts. This means that courts of first instance are bound by the decisions of more senior

courts on points of law. This ensures that the interpretation of law is consistent and (consequently) understandable. The administrative proposals will essentially allow decision makers to apply the law or rules arbitrarily.

(e) **They lack speed** – The courts, when required, are able to resolve the issues before them quickly. The Standards System tends to take months if not years to reach a decision.

(f) **They lack transparency** – Unlike the courts, which are public, the current standards system involves evidence and decision making behind closed doors. In Westminster, the Public Accounts Committee is conducting an inquiry into the use of arm's length regulatory bodies because, in the words of its Chair, we should be "alarmed by the lack of accountability and transparency" of many arm's length bodies.[11]

Does the ICDR Model impact on Senedd privilege?

51. It has been suggested that the proposal will impact on parliamentary privilege. It should be noted that:

(a) The Senedd "privilege", as it currently stands, will not be impacted by the ICDR Model (or any equivalent proposal) because it does not protect MS from legal action save for defamation and contempt of court.

(b) Even if Article IX of the Bill of Rights is extended to the Senedd, that provision does not give politicians the right to make false or misleading statements. Both Westminster and Senedd politicians are already bound by rules which prohibit misleading statements. The current measures under discussion simply concern how those existing duties are enforced. It is within the Senedd's power to waive that privilege in cases of deliberate deception.

[11] https://www.ft.com/content/897c7451-b615-4097-8a80-fbae505b529d (last accessed 24 November 2024)

Is the "real possibility" test appropriate?

52. The ICDR Model envisaged the use of the "real prospect" test (which is very much a "recognised legal term of art")[12] rather than the "real possibility" test. The drafting should be corrected accordingly (and has been in Version 2).

Does the White Paper contain misleading statements?

53. We have not been able to identify any misleading statements in the ICDR White Paper (save for those which were corrected in Version 2). The paper is fully referenced, and its veracity can be confirmed from those references. Where new information has required updates, these have been included in Version 2 of the ICDR White Paper. It is important to distinguish between statements which are "false or misleading" and statements which contain differences of opinion. The courts are experienced at distinguishing between statements of fact and opinion because they are already required to by section 3 of the Defamation Act 2013. The ICDR Model accounts for this by limiting its application purely to statements of fact. We have adjusted the drafting of the model clauses to make it clear that the regime should not apply to statements of opinion.

[12] See, for example, ***Mansell v Tonbridge and Malling Borough Council*** **[2017] EWCA Civ 1314.**

Pencadlys yr Heddlu | Police Headquarters
Glan-y-Don, Colwyn Bay, LL29 8AQ

11th December 2024

Dear Hannah,

I write in response to your letter dated the 2nd of December 2024 following my attendance at the Standards of Conduct committee.

The Police support Returning officers and the management of elections in Wales. The resources usually come from each forces Economic Crime Unit (ECU) as they have the required expertise and knowledge to deal with such incidents.

Each force will provide a Detective Sergeant or above to undertake the role of election single point of contact (SPOC). Training for the role is provided by the City of London Crime Academy election offences training.

Once suitably trained the relevant Officers have responsibility for the following:

- Engaging with the Electoral Commission for Wales and the electoral service manager.
- Engaging with other Police SPOC to ensure a consistent approach across Wales.
- Engaging with local authorities to identify any communities or locations vulnerable to electoral malpractice.
- Work with Tarian to provide Cyber protect advice to relevant stake holders in advance of any election.
- Provide briefings to all candidates, agents and police personnel to outline relevant legislation as well as advice on how to manage appropriate contact during an election period.
- Investigate each offence reported to police and any subsequent offences that may be uncovered during the initial report.

SPOC's attend a number of meetings in the several months leading to any election. These officers will see a significant increase in workload 2 months prior to any election. Within this time, they spend a minimum of 4-5 hours a day dedicated to this role in addition to managing their usual role. This is purely around preparation.

Any additional investigations that arise such as complaints or appeals will be managed by other members of staff. Quite often due to the public profile of the individuals involved there is a requirement

for senior officers to have responsibility for these cases. This is to ensure the investigation progresses in an effective and timely manner.

Below is a breakdown of the number of complaints made to police across Wales and the subsequent outcomes whether they be caution or prosecution.

Date	No of Complaints	Caution	Prosecution
2008	5	1	0
2010	2	0	0
2011	2	0	0
2012	9	0	0
2014	2	0	0
2015	6	0	0
2016	4	0	0
2017	29	2	0
2018	4	0	0
2019	6	0	1
2020	8	0	0
2021	1	0	0
2022	6	0	0
2023	2	0	0
2024	25	1	0

It should be highlighted that the above chart does not include any outcomes whereby words of advice has been given. In addition, the figures do not accurately reflect other demand that will arise from such investigations. These will include support from the force's communication teams and any times reviewing challenges in relation to outcomes.

I hope that this provides the additional information you were looking for.

Yours sincerely

Amanda Blakeman KPM
Chief Constable
North Wales Police

Document 3 - Elkan Abrahamson

Submissions to the Standards of Conduct Committee of the Welsh Senedd as requested in email dated 9/12/24.

A. Scope

I have been asked to comment on 3 options under consideration by the Senedd Standards of Conduct Committee for dealing with individual member accountability:

'1. The creation of a criminal offence of deception which would be investigated by the police and tried before the Criminal courts.

2. Using an existing investigative body such as the Public Services Ombudsman and an independent Welsh tribunal, such as the Adjudication Board for Wales (if the making of false or deceptive statements of fact is to be a matter for civil sanction).

3. Strengthening the existing Code of Conduct (rule 2) ….to prohibit more explicitly willful lying or deception and strengthen the sanctions which could be applied. These would continue to be dealt with through the existing mechanisms of the Senedd Commissioner for Standards and considered by the Standards of Conduct Committee and Senedd.'

I was also invited to consider any other proposals.

I note that the email to me suggested that the terms of reference of the committee were restricted to considering mechanisms for disqualification of Members and candidates. No mention was made of other sanctions (criminal or civil) although option 1 refers to a criminal offence and option 2 refers to civil sanctions. The discussions before the Committee to date have included consideration of criminal sanctions and it is unclear to me whether this possibility has now been excluded. I discuss (and support) it but accept this may be beyond the scope of my invitation.

B. Personal Background.

I am most grateful to be given this opportunity and I should start by saying that I do not profess to have any expertise in the Welsh system of governance. I have considered the submissions and oral testimony to date, am mindful of the task which the Senedd has taken on and make my comments with what I hope is appropriate humility.

I am a qualified solicitor specializing in major inquests and Inquiries., a non- executive director of an NHS trust in England and a director of Hillsborough Law Now (HLN), a pressure group set up to introduce among other things a duty of candour for public officials enforceable through the criminal process. I am also part of the lawyer's drafting group for Whistleblower UK. I have represented bereaved families in the Hillsborough inquests, the Birmingham pub bombings inquests, the Manchester Arena Inquiry and currently the Covid Inquiry. The views expressed here are however my own.

C. Hillsborough Law

It might be helpful to set out the history of the Hillsborough Law campaign and why HLN considers that a criminal process with criminal sanctions is necessary.

The Hillsborough Disaster occurred during a football match between Liverpool FC and Nottingham Forest on 15 April 1989. An influx of Liverpool fans into the standing-only stalls in the Leppings Lane stand of Sheffield's Hillsborough Stadium caused overcrowding of the pens. This overcrowding resulted in 97 deaths and 766 injuries – the highest death toll in British sporting history.

The 1990 Taylor Inquiry Report on the incident found that the police had neglected their duty by opening all of the gates in an attempt to ease overcrowding, subsequently leading to the crush in the pens as fans rushed in. The police subsequently fed the press stories which were found by the inquest to be false. They blamed the crush on Liverpool supporters – alleging it was due to their hooliganism and drunkenness. The impact of this has persisted even after Lord Justice Taylor's report, which found no fault on the side of the supporters, and failure of control by South Yorkshire Police. A 2009 Independent Panel resulted in similar findings, and a second coroner's inquest in 2016 found the 96 (now 97) supporters were unlawfully killed due to gross negligence by the police and ambulance services. However, besides a single conviction of the stadium's health and safety officer (resulting in a £6500 fine), all prosecutions have failed..

It became clear to us that a statutory duty of candour combined with criminal sanctions for failure to comply was the only effective mechanism for ensuring that public servants tell the truth to inquiries.

Bishop Jones was asked to write a review following the Hillsborough Inquests and this – 'The patronising Disposition of Unaccountable Power' -was laid before the UK parliament in November 2017. The bishop included Hillsborough Law as drafted by us in his report and recommended that once the Law Commission finished its work on the offence of Misconduct in Public Office (now finished) consideration should be given to the Bill. The Bill (formal title: The Public Authority (Accountability) Bill) is now under consideration. The bishop also recommended that a Charter for bereaved families should be adopted by leaders of public bodies. That Charter – accepting the need for openness – has not made a shred of difference that I can see to the conduct of public bodies. This makes the point that only legislation with effective sanctions can actually change the position.

D. The failure of the NHS duty of candour.

HLN was always of the view that while codes of conduct may have a role to play in, they will not succeed in changing behaviour unless backed up by a system of enforcement which includes effective sanctions.

I emphasis 'effective' and in this respect would mention the failure of the NHS duty of Candour. This requires some elaboration.

In response to the Mid Staffordshire Inquiry, a duty of candour was imposed on the NHS in The Health and Social Care 2008 (Regulated Activities) 2014. It was doomed to failure from the start. It requires that:

> 'As soon as reasonably practicable after becoming aware that a notifiable safety incident has occurred a health service body must—
>
> (a)
>
> notify the relevant person that the incident has occurred in accordance with paragraph (3), and
> (b)
> provide reasonable support to the relevant person in relation to the incident, including when giving such notification.'

Its main weakness is that it imposes a duty on organisations not individuals. Furthermore, even when an offence is proved the penalty imposed on the organisation (again, not an individual) is minimal - a level 4 fine, with a current maximum of £2500.

There have been a few, very few, prosecutions since the duty of candour was introduced.

The first reported prosecution which went to court was when a Mrs. W suffered a perforated oesophagus during an endoscopy in December 2017.

The procedure was abandoned, and Mrs. W was transferred to a ward for observations. She collapsed on the ward and later died.

Under the regulations, Mrs. W's family were entitled to a full explanation and a prompt apology. Unfortunately, Mrs. W's family received neither a prompt apology nor a full explanation regarding the tragic events that took place prior to her death.

CQC brought a prosecution against University Hospitals Plymouth NHS Trust for their failure to disclose the details of how she died and the failure to apologise within a reasonable timeframe.

The Trust admitted guilt and was fined £1600 by Plymouth Magistrates Court., plus a £120 victim surcharge and £10,845.43 in court costs.

Prior to that Bradford Teaching Hospital in January 2019 was fined £1,250 for failing to apologise to the family within a reasonable period of time, following the death of their baby

boy in July 2016. There also appears to have been a prosecution resulting in the Royal Cornwall Hospitals NHS Trust being fined £16,250 for 13 fixed penalty notices, again for failing to comply with the duty of candour.

Notably absent from this list are the many NHS scandals leading to inquiries.

In 2020 the CQC published a report into maternity services at the Princess Royal hospital and said that the Trust must 'ensure grading of incidents reflects the level of harm, to make sure the duty of candour is carried out as soon as possible.'. The Princess Royal hospital is part of the Shrewsbury and Telford Hospital NHS Trust and of course featured heavily in the Ockenden Inquiry into their maternity services. This report found many instances of a failure to observe the duty of candour - investigating without notifying the mother (page 60) ; not following the duty of candour following a failed forceps delivery (and subsequently failing to train the junior doctor who had not complied with his duty) (P.109) - see the Shrewsbury and Telford Maternity services Inquiry .

Further examples of NHS scandals abound, the ongoing Letby Inquiry being the most recent. They all have in common a theme of issues being raised and ignored by senior management despite the statutory duty of candour.

The conclusion I draw is that a voluntary code of conduct is ineffective as is any regime, statutory or otherwise, which has no effective 'bite'.

The only regime likely to be effective is one which has sufficiently grave sanctions as to act as a deterrent. What those sanctions need to be will inevitably vary from case to case.
The Senedd initiative comes against a background of mistrust of politicians (as discussed in much of the oral testimony to this committee), of politicians and candidates misleading the public and a culture of what might be termed 'misleadership'.
There is an urgent need to restore trust in the democratic process (again this has been discussed in committee) which in turn requires the restoration (or perhaps introduction) of integrity into public service.
No single measure will achieve this and any consideration of sanctions for MS's should take this into account.
We need and expect in addition a statutory duty of candour for public officials. We need a whistleblowing regime which will protect whistleblowers in public life (and in this respect `I would refer you to the website of Whistleblower UK - WBUK.com) . Finally, a review of the effectiveness of the Nolan criteria is due - I would hope the criteria could be tightened up and backed up by a more effective sanctions system.

E. The 3 options
Turning now to the 3 options mentioned at the top of these submissions:
Option 3 -in effect a process relying on the Senedd and/or one of its committees - has the disadvantage that the range of sanctions is limited, and the process is internal. It seems unlikely that this will assuage public concerns about integrity. It also carries the substantial risk of becoming politicised. In this respect I note that the Guardian on 25/12 reported that the Parliamentary Modernising Committee is to recommend that complaints about MP's conduct relating to bullying harassment and sexual misconduct be referred to the Independent Complaints and Grievance system (ICGS) instead of letting political parties deal with them. (See article by Eleni Courea, The Guardian 25/12/24).

Option 1 - The most effective system for dealing with lies by MS's and candidates would be the criminal process. There is no doubt that the prospect of imprisonment focusses the mind. However, it is considered by some that using the criminal process has implications for separation of powers and there are other reservations such as the floodgates argument. I do not believe there is a separation of powers issue here. The Senedd has the power to regulate the conduct of its members in any way it sees fit. I do recommend and recognize that what S' say in the Senedd should (as in the UK parliament) be fully protected and this

is what I understand to be parliamentary privilege. This protection is essential to enable MSs to raise injustices on behalf of their constituents without fear of reprisal. I would suggest this is accompanied by the responsibility to filter out wild and speculative accusations but that is another matter.
. Hillsborough Law Now considered whether to extend the duty of candour to politicians. We decided this was not feasible because there was no prospect of garnering political support (among MP's) for this, but we were also conscious of the constitutional difficulties. I do not think it constitutionally impossible to fashion a criminal offence although consideration needs to be given to the implications and the sanctions.
I have seen argument before you that using Judges will help reassure the public as Judges are trusted more than MP's. That may turn out to be flawed. Judges enjoy public trust to some extent because they are largely separate from the political process – the more Judges are involved in it the more the possible threat to their perceived independence and thereby the public trust in them. Ultimately the primary aim is not to reassure the public, it is to restore integrity to public life. The reassurance should follow from that.

There is a distinction to be drawn here between lying by MS's and lying by political candidates. Candidates do not have the protection of parliamentary privilege, nor do they have the constitutional protection of being part of the legislature. Indeed, the committee has heard from Chief Constable Amanda Blakeman that s.106 of the Representation of the People Act 1983 creates a criminal offence of making a false statement while a candidate. I would suggest that candidates can and should be subject to a standard criminal process with such sanctions as the government considers appropriate (including imprisonment). I would suggest that the objections to criminal process for candidates can all be surmounted. The 'floodgates' argument is a red herring -there is no evidence that there would be a flood of prosecutions or even complaints and the police are well able to deal with frivolous/vexatious complaints. Objections have been raised that proving the elements of a criminal offence beyond reasonable doubt would be difficult, but that is as it should be. It would be preferable for any offence to be triable only by a judge and jury as this would involve trial by peers.
Once someone is elected by a democratic process it is still possible to justify a criminal process as the authority for that will come from the legislation passed by the Senedd itself.

Option 2 - Using an independent body and empowering it to impose civil sanctions.
The effectiveness of this depends on the quality of the process and the sanctions which the body could impose. Would it be empowered to suspend/disqualify a sitting MS? Would it have the ability to act as a judicial or quasi-judicial body? Would the alleged miscreant be allowed legal representation? Would there be a prosecutor equivalent with adequate resources to investigate? Would civil sanctions ensure compliance? Any MS with a substantial backer could happily accept a fine. Finally, as raised in oral hearing before you, would it send the right message to the public?

In conclusion I would suggest that Option 3 is inappropriate and ineffective, Option 2 has limitations and is unlikely to be successful and Option 1 – a criminal process ensuring full rights for the accused MS/candidate – is the most sensible and appropriate of the three suggested.

I thank the Committee for their invitation to submit and would be happy to expand on any of these points if thought helpful.

Elkan Abrahamson
Director
Broudie Jackson Canter solicitors

3/1/2025

Hannah Blythyn MS
Chair of the Standards of Conduct Committee
Welsh Parliament
Cardiff Bay, Cardiff, CF99 1SN

12 December 2024

Individual Member Accountability Inquiry

Dear Hannah,

Thank you for your email following Full Fact's evidence session at the Individual Member Accountability Inquiry and the request for further information. We've outlined responses to your questions below, but please don't hesitate to reach out if you require further information.

Full Fact's position on the draft Public Authority (Accountability) Bill, currently making its passage through the UK Parliament, and any parallels this may have with this Committee's inquiry into introducing the offence of deception as a criminal or civil offence.

A draft of the Public Authority (Accountability) Bill, also known as the Hillsborough Law, has not yet been tabled, so we are mindful to not put a position on record without having reviewed the full text of the Bill. However, from what has already been shared and from previous versions of the Bill, we are supportive of initiatives that increase honesty and transparency in order to support inquiries and investigations.

Both the definition of the 'duty of candor' and how deception is defined within any future legislation proposed by this inquiry will need to be carefully considered.

The correction system currently in use in the UK Parliament, and any data you may be able to share in relation to the frequency of issuing correction notices and how often they lead to corrections.

Fact checks alone are not enough to halt the spread of inaccurate information, however what makes Full Fact different from other fact checking organisations is when we see an inaccurate claim being repeated, we get in touch with those responsible to correct the record. Evidence tells us that corrections can be most effective if whoever said the claim corrects it themselves. This helps us to change attitudes and behaviours, encourage a culture of accuracy, and gather evidence about how well the systems meant to stop bad information reaching the public are working.

In our 2024 report we outlined what steps needed to be taken to ensure MPs in Westminster had the same ability to correct the record as Ministers. The recommendations of the Procedure Committee were implemented in April 2024, improving the corrections system to allow MPs to correct the record by making a point of order and writing directly to Hansard.

The full effectiveness of the correction system can't be measured by Full Fact alone. Given it is not possible for us to monitor and fact check all claims made by politicians, it is not accurate to draw

comparisons about the overall behaviour of individual MPs or political parties. However, our work has shown that politicians are willing to correct the record when inaccurate statements are pointed out to them.

We've secured corrections from former Prime Ministers, including David Cameron and Rishi Sunak, both sides of the House of Commons, national charities, and every major newspaper. All of this is outlined on our website in the 'Your MP' section, which is updated every time we intervene after a fact check. However, some MPs also refuse to correct the record and Full Fact alone cannot force them to do so.

As we presented in the session, we would support the committee exploring what parliamentary bodies could encourage and enforce correction notices. This work also should also be done in parallel with greater training and guidance from political parties to ensure that MPs are willing to take on responsibility for their statements.

I wish you the best with the rest of the inquiry and we look forward to reading what the committee recommends. Restoring trust in politics is an issue of great importance to us all and we are glad it is being given due consideration through the committee.

Best wishes,

Azzurra Moores
Policy Lead, Full Fact

STANDARDS OF CONDUCT COMMITTEE

BILL ON POLITICAL DECEIT

Submission

Professor Conor Gearty FBA, KC (Hon)

LSE and Matrix Chambers

Introduction

1. I support this legislation. I believe that it is well-thought out, carefully drafted and addresses a need of immense importance. That need is the urgent imperative of effectively addressing the threat to our system of representative democracy that is posed by the deliberate deployment of deceit by malign political actors.

The Wider Picture

2. Our parliamentary system is based on three assumptions: (i) that those elected to represent constituencies in a democratic assembly will deploy their best efforts (as they see them) to improve the lives of the constituents whose interests they have been elected to represent; (ii) that political decision-making results from rational discussion; and (iii) that all political actors engage in that rational discussion in good faith, and so neither seek to promote hidden agendas nor advance covert external interests. None of these three assumptions can be taken for granted. This has long been accepted with there being rules in place to prevent hidden conflicts of interests or the taking of bribes.
3. The time has now come to act against a new kind of public wrong, the use of deliberate deceit to achieve control of power in democratic systems. There has always been a tendency to lie to achieve political ends. The issue has become acute because changes in technological capacity have made lies easier plausibly to assert and then to maintain, with evidence mustered to show their falsity not being guaranteed to prevail. These lies can and do then play a critical role in influencing voter choice, another effect of recent technological change. Scattered everywhere, they stick to the political ground, changing the nature of political discourse.
4. In Wales as in the rest of the United Kingdom, no malign political actor has as yet achieved power through deployment of lies in the pursuit of power. It has however happened elsewhere and could come to these islands in the nearish future. A type of potential legislator has come to the fore who has shown themselves able and willing to deploy lies to further their political agendas, and

to use technological change to achieve their ends. Such persons are often well-funded and so heavily resourced.

5. The time has come to act against this kind of malign influence before it is too late. Representative democracy has experience of resistance to an 'enemy within', demonstrated by its successful defence of itself against Communist subversion during the middle years of the last century. It is time to dust down the old democratic survival toolbox and find within it what might be relevant to today.

The Proposed Bill

6. The measure being proposed goes some way towards meeting the challenge of deceit identified above. I strongly support it.

7. So far as the detail is concerned, I have the following comments and suggestions.
   i. The wrong at which the Bill is aimed should be processed through the criminal rather than the civil law. The criminal law has an important symbolic role as a signifier of grievous wrong. The wrong here is not only or even necessarily to individuals; it is to the system as a whole. We need to adjust our thinking to see such wrongs as deeply serious: the deployment of the criminal law will play a part in that.
   ii. The criminal enforcement of these provisions must involve non-judicial actors, analogous to a jury. A magistrate who determines there is a case to be heard should be able to draw ten residents from the electoral register willing and able to participate in the relevant proceedings and to give judgment. It must be through the judgment of the 'ordinary' citizen that such liars come to be punished.
   iii. Of course there are issues of speed and efficiency which would point towards judge-only trials. But there is a greater interest, in my view, in engaging the community in the assessment of wrong here, and in doing so hopefully heading off any assertion that such proceedings are merely an elite kind of 'lawfare' (or at least rendering such claims less plausible).
   iv. These criminal proceedings should be televised, and the authorities should be given special powers to publicise such proceedings in ways that they see fit and which are designed to maximise the impact of such proceedings on the wider public. They should be short, and designed in a way that engages a watching public. The jury should be able to engage the defendant in examination during the trial. Consideration should also be given to televising the deliberations of the judge and jury (with the magistrate acting as a kind of foreman).

v. Where it culminates in a guilty verdict the process will often be the punishment in itself, but formal publishment should also be available to the sentencing judge, including banishment from public life.
vi. Consideration should be given to permitting the jury to engage with the judge in consideration of not just guilt but the appropriate sentence as well.
vii. There should be no prohibition on discussion by these court participants of the process that led to their decision.

Building on this reform

8. The Bill is a very god start, but more can be done.

9. Each democratic system should have an Office for Democratic Integrity (ODR), responsible for the preservation from attack of our system of representative government. As well as the proposed law on deceit, all the current laws and practices related to the health of the democratic system should be brought under its aegis, and (in the UK context) devolved and local offices of the ODR should also be established.

10. Insofar as for example the public service ombudsman and the adjudication board for Wales have such responsibilities then these should be transferred to the ODR.

11. As indicated above the ODR should have responsibility for, among other functions, the detection and punishment of deliberate lying in the pursuit or exercise of power, fulfilling a CPS-type role in this area. This is the wider (and it is acknowledged) more ambitious context in which to set this Bill, a trail-blazer for an urgent and altogether much wider engagement in the defence of our system of government. Wales can set the pace here for defence of the parliamentary system of government – before it is too late, not just in Wales but everywhere.

Conor Gearty

London

6 January 2025

Document 6 - Henrietta Catley

Towards a Criminal Offence for Politicians Making Deceptive Representations to the Public

Abstract

In recent years politicians have been increasingly accused of making deceptive representations to the public. The false or misleading information imparted in these representations has the potential to influence how the public forms political preferences, leading to broader democratic implications when the public then engages in democratic procedures. These developments have exposed the need to reflect on what is and should be being done to deter politicians from these representations. My purpose in this paper is two-fold. I explore why these representations are problematic and set out a case for why legal reform is necessary. In doing so I make the case for a new legal solution. Specifically, I advocate for a new criminal offence to address the most egregious deceptive political representations which are made to the public.

Introduction

Deceptive representations (material statements of fact which are knowingly or recklessly false or misleading) have always been pervasive in the political domain.[1] Unlike other types of deceptive statements (like opinions or vague comments), deceptive representations carry greater credibility and influence. They are made with a sincerity and certainty which offer the public something to rely on and induce them to have false beliefs.

As a public we are used to our politicians making these representations. In the 1990's John Major lied to the public about engaging in peace talks with the IRA,[2] and just over a decade later Tony Blair claimed that there was intelligence which showed[3] that Iraq had weapons of mass destruction when this was not the case.[4] Yet recent shifts have meant the situation is now more poignant than ever

[1] P. Bernal, *The Internet, Warts and All* (Cambridge University Press, 2018) pp.229, 234-239, 241.
[2] A. Bevins, E. Mallie and M. Holland, "Major's secret links with IRA leadership revealed" (28 November 1993) *The Guardian, https://www.theguardian.com/uk/1993/nov/28/northernireland*
[3] BBC, "Breakfast with Frost" (26 January 2003) *BBC, https://www.bbc.co.uk/programmes/p00pnb0d* at 5:42-6:10.
[4] Committee of Privy Counsellors, *The Report of the Iraq Inquiry: Executive Summary* (HC 2016, 264) paras 540, 796. See also Gordon Corera, "How the search for Iraq's secret weapons fell apart" (13 March 2023) *BBC, https://www.bbc.co.uk/news/world-64914542*

before. Greater scrutinisation from democratic watchdogs has uncovered more examples of deception, thereby exposing the extent to which politicians are trying to deceive the public. Just a few years ago, Boris Johnson's claim that the UK will make a gain of £350 million a week if it leaves the EU became a notorious example of deception in politics.[5] The implication from the claim was that this was a net gain, when in fact it "did not take into account the rebate or other flows from the EU to the UK public sector […]".[6] Another prolific instance was Boris Johnson's claim that he did not participate in Covid-19 breaches, saying "anybody who thinks I was knowingly going to parties that were breaking lockdown rules […] [are] out of their mind". This is despite having been previously reported as saying "this is the most un-socially distanced party in the UK right now […]".[7] The point is that greater scrutiny has brought to light the regularity with which these representations are made and the extent to which our politicians are trying to deceive us. All this creates an opportunity for reflection. In particular, it invites questions about what the UK's approach to addressing deceptive representations is, and whether it is up to par.

Now, the UK has a multitude of mechanisms at its disposal to recognise and sanction deceptive representations. The problem is that these mechanisms are deficient in various ways, either in terms of not being utilised, having a lack of applicability or being otherwise fundamentally flawed. The result is that politicians are not discouraged from making these representations. In the meantime, the false or misleading information being imparted has the potential to influence how the public forms political preferences. This can have broader democratic implications when the public engages in democratic procedures.

What I argue for is a new criminal offence to complement the existing framework. The criminalisation route has been attempted before, as with the Elected Representative (Prohibition of

---

[5] A. Asthana, "Boris Johnson: we will still claw back £350m a week after Brexit" (16 September 2017) *The Guardian, https://www.theguardian.com/politics/2017/sep/15/boris-johnson-we-will-claw-back-350m-a-week-post-brexit-after-all*

[6] Office for National Statistics, "Leave campaign claims during Brexit Debate" (7 February 2017) *ONS*, *https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/leavecampaignclaimsduringbrexitdebate*

[7] S. Ravikumar, K. MacLellan and W. James, "UK's Boris Johnson and the 'partygate' scandal" (15 June 2023) *Reuters, https://www.reuters.com/world/uk/uks-boris-johnson-partygate-scandal-2023-06-15/*

Deception) Bills of 2006 and 2023. However, these attempts have been unsuccessful and never come close to being enacted into law. My suggestion is that we should still use the criminal law but take a different tact and try to make it more appealing. I advocate for a new and more narrowly drafted criminal offence (compared to what has previously been suggested). Specifically, one which is geared towards the most egregious deceptive representations which are made to the public. Importantly, I argue that this strikes the balance between intervening to provide some protection from democratic harms whilst not overextending the criminal law or giving substance to frequently cited objections.

To substantiate this argument, I adopt a tripartite structure. I begin by using political theory to explain why making deceptive representations to the public is problematic. I then critically assess the efficacy of the mechanisms in the UK's current approach, finding it to be deficient. In the final part of the paper, I propose a way that this deficiency can be addressed and advocate for a new offence. It is beyond what can be achieved in this paper to outline a complete draft offence, but I do put forward a provisional idea of what it could involve. In particular, I highlight certain elements which should form the foundation of the offence, with the aim of capturing the most egregious deceptive representations.

The Problem of Deceptive Political Representations

Why are deceptive representations to the public problematic?

Understanding why deceptive representations are a problem returns to the broader question of how the public forms political preferences. Elite influence as a whole is widely appreciated as being a significant factor in shaping political opinions. "Citizens have clear incentives to take political cues from those more knowledgeable, typically experts or elites whose views are conveyed by the media".[8] The elites which carry this influence are a "wide range of individuals and organizations, including politicians, political officials, policy experts, interest groups, religious leaders, and journalists".[9] Politicians in particular (by this I mean those who are in Parliament or Government) are fundamental

[8] M. Gilena and N. Murakawa, "Elite Cues and Political Decision Making" (2002) 6 Political Decision Making, Deliberation and Participation 15, 15.
[9] Gilena and Murakawa, "Elite Cues and Political Decision Making" (2002) 6 Political Decision Making, Deliberation and Participation 15, 16.

to forming and shaping public opinion. They have the unique position of not only being professionals in the political field but also being informed of and party to the inner workings of policy. As put by Christiano, "the evaluation of policy includes many different elements and areas of expertise" e.g., knowledge of: the sciences, law and policy, how to compromise in the political environment, and local knowledge.[10] This is something which the majority of the public is not well-versed in. The ideal is that the public draws on the expertise of those who are: politicians.

Deference to politicians is similar to how you treat any area outside of your expertise. When you are not an expert in the field you defer to someone who is- this is a logical form of specialist labour-division. In the context of politics, the public subcontracts expertise in policy to politicians with the expectation that relevant and truthful information is relayed back. The public then draws on and evaluates this information, using the evidence and the soundness of an argument's logic to form their own opinions.[11] The issue is that when politicians deceive this arrangement is threatened because the information being imparted is corrupted. The public then potentially forms political opinions from a knowledge base which is comprised of false or misleading information.[12]

With that being said how the public forms political opinions is a multi-faceted issue. In particular, heuristic factors (cognitive short-cuts which enable the public to "be knowledgeable in their reasoning about political choices without necessarily possessing a large body of knowledge about politics")[13] are increasingly recognised as important. This school of thought evolved from the Downsian economic view of public reasoning. In Downs's view, acquiring and evaluating information is a costly activity[14] - requiring a considerable amount of time and effort which most rational people are not

[10] T. Christiano, 'Rational deliberation among experts and citizens' in
J. Parkinson and J. Mansbridge (eds) *Deliberative Systems* (Cambridge University Press 2012) p.28.
[11] Gilena and Murakawa, "Elite Cues and Political Decision Making" (2002) 6 Political Decision Making, Deliberation and Participation 15, 17.
[12] J. Oppenheimer, *Principles of Politics* (Cambridge University Press 2012) p.15.
[13] P.M. Sniderman, R. A. Brody, and P.E. Tetlock, *Reasoning and Choice: Explorations in Political Psychology* (Cambridge University Press 1991) p.19.
[14] A. Downs, *An Economic Theory of Democracy* (Harper & Row 1967) p.207.

willing to expend. [15] The idea is that instead they turn to other more efficient resource-saving devices.[16]

Since Downs, a number of scholars have analysed how the public actually forms political preferences, and what has ensued is a growing recognition of heuristic influences.[17] LePoutre for instance, puts great emphasis on the influence of social group membership. He argues that there is an "assum[ption in democratic theory] that […] [the public is] trying to form accurate or reliable political judgments. But there is ample evidence suggesting that, when it comes to politics, people simply accept whatever their social group tells them to believe",[18] through intuitive[19] (social or emotional) reasoning.[20] The core idea is a social group is partly defined through exposure to a "distinctive set of shared social constraints and enablement's by the laws, norms, and physical infrastructure that constitute the social context".[21] The by-product is a shared social perspective with shared constraints and experiences at the forefront. LePoutre argues that this perspective can be influential when group members look to form political preferences. Members take a cognitive shortcut, trusting and using the opinion of other more informed members as a cue for their own political preferences. This is on the assumption that most members will share certain priorities.[22]

The public also draws upon political partisanship to guide the publics preferences. Say, for instance, that you are affiliated with a particular political party and the party approves of a particular immigration policy. You (as a member of the public) are uninformed about the policy and have not yet formed an opinion. To shape your opinion, you refer to the stance your party takes, placing trust in it because you assume that it will have a similar outlook and values to your own.[23]As Kuklinski and

---

[15] Downs, *An Economic Theory of Democracy* p.42.
[16] J. Mondak, "PUBLIC OPINION AND HEURISTIC PROCESSING OF SOURCE CUES" (1993) 15(2) Political Behaviour 167, 168.
[17] E.g. Achen and Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* p.272 and I. Somin, *Democracy and Political Ignorance: Why Smaller Government is Smarter* (2nd edn, Stanford University Press, 2016) pp.62-79
[18] M. LEPOUTRE, "Democratic Group Cognition" (2020) 48(1) Philos. Public. Aff. 40, 41
[19] M. LEPOUTRE, "Democratic Group Cognition" (2020) 48(1) Philos. Public. Aff. 40, 41, 43, 51.
[20] Achen and Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* p.232.
[21] LEPOUTRE, "Democratic Group Cognition" (2020) 48(1) Philos. Public. Aff. 40, 48.
[22] LEPOUTRE, "Democratic Group Cognition" (2020) 48(1) Philos. Public. Aff. 40, 50-51.
[23] J. Mondak, "PUBLIC OPINION AND HEURISTIC PROCESSING OF SOURCE CUES" (1993) 15(2) Political Behaviour 167, 182-184.

Quirk note, "[b]y merely attending to party labels, voters can compensate for a lack of reliable information […]".[24] The classic example of using party affiliation to act as an opinion cue is when people are called to vote on political candidates. Often the public tends not to explore the policies of a particular candidate. Instead, they will often look to the party the candidate is affiliated with to decide how to vote, assuming that the candidate will have a particular outlook because they have a certain party-membership.

Aside from social or political cues, scholars are increasingly noting the use of gut instinct or emotive influence as another heuristic method. Popkin in his pioneering work on voting preferences in the US, places emphasis on the likeability of a party or candidate, as well as the inferences which can be made about their character e.g., their sincerity.[25] These inferences can then be used by the public to influence how they form a political opinion e.g., whether a policy is viewed favourably. Seen in this light, political opinion formation needs to be recognised as a complex and multi-faceted theory.

Whilst it is important to stress that this is not a straightforward issue and that a politician's influence may not be determinative in how opinions are formed, it still has a role.[26] Even new empirically-based studies note its significance. Take, for instance, Clarke et al's mass observation project into voting in the Brexit referendum.[27] The results indicated that when developing opinions, panellists often fell back on feelings.[28] The findings showed that '[m]any panellists looked to the campaign for help, at least initially. They read leaflets and newspapers, watched television and listened to the radio'.[29] The problem was that the misleading, unverified and contradicting claims left them feeling uninformed,

---

[24] J. H. Kuklinski and P. J. Quirk, "Reconsidering the Rational Public: Cognition, Heuristics, and Mass Opinion" in *Elements of Reason Cognition, Choice, and the Bounds of Rationality* (Cambridge University Press 2012) p.155.
[25] S.L. Popkin, *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (University of Chicago Press 2020) pp.7, 44, 65. See also, J. H. Kuklinski P. J. Quirk, "Reconsidering the Rational Public: Cognition, Heuristics, and Mass Opinion" in *Elements of Reason Cognition, Choice, and the Bounds of Rationality* (Cambridge University Press 2012) pp.153-159.
[26] G. Markus, "The Impact of Personal and National Economic Conditions on the Presidential Vote: A Pooled Cross-Sectional Analysis" (1988) 32 American Journal of Political Science 137, 137–154.
[27] N. Clarke, W. Jennings, J. Moss and G. Stoker, "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 110 -118.
[28] Clarke, Jennings, Moss and Stoker., "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 107.
[29] Clarke, Jennings, Moss and Stoker, "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 113.

and the public was forced to use their instinct to fill that deficit.[30] Another illustrative example is Arceneaux's cross-national analysis of the impact of election campaigns as informational tools and their use by voters.[31] Survey data from 9 European countries across 12 years, indicated that voter learning actually improved when party's campaigned. Although Arceneaux did observe that "voters' pre-existing attitudes and assessment of things beyond the control of political campaigns, like the economy, have the strongest impact on voting decisions, [he also noted that] campaigns play a major role in producing these effects".[32] In other words, whilst heuristic factors were present, their existence did not invalidate the fact that information being imparted by politicians also had influence.

If we follow this reasoning, then we can also hold that the deceptive representations being made by politicians is problematic. It is giving the public a flawed knowledge base on which to make decisions. This is something which has broader democratic implications when the public then engages in democratic procedures whether it be institutional e.g., voting in referendums or elections, or alternative and ongoing signalling procedures e.g., protesting, petitioning or the critiquing of policy decisions.[33] It is on the basis that deceptive political representations play a part in political opinion formation and potentially give rise to democratic implications, that I suggest that this issue is problematic. Having explored why this is a problem, I will now move onto the next part of this paper in which I critically assess the UK's current approach.

The Failure of the Current Approach

The current approach to tackling these representations can be broadly distinguished into two themes: an approach based on a lack of intervention (i.e. the self-correcting public debate) and formal

---

[30] Clarke, Jennings, Moss and Stoker, "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 115-116.

[31] K. Arceneaux, "Do Campaigns Help Voters Learn? A Cross-National Analysis" (2005) 36 British Journal of Political Studies 159, 160, 164-169.

[32] Arceneaux, "Do Campaigns Help Voters Learn? A Cross-National Analysis" (2005) 36 British Journal of Political Studies 159, 160.

[33] See S. Marien and H. Serup Christensen, "Trust and Openness: Prerequisites for Democratic Engagement?" in K. N Demetriou (ed) *Democracy in Transition Political Participation in the European Union* (Springer 2013) 109, for discussion on alternative forms of activism. See more generally, R. Inglehart, *Modernization and postmodernization: Cultural, economic, and political change in 43 Societies* (Princeton University Press 1997) p.307.

regulation.

Self-correcting public debate

The first part of the approach is based on non-intervention, essentially, relying on the idea that public discussion has the ability to self-correct. Such a theory is advocated by a number of free speech scholars including Mill,[34] Brandeis and Holmes,[35] and Mieklejohn,[36] who have all suggested that the public debate has the capacity to identify problematic speech (e.g., false or dangerous rhetoric) and remove it from circulation. The theory is that unconstrained public discussion provides an open platform for ideas, so initially all ideas have equal status. However, as time and the discussion develop, the public is able to scrutinise and challenge different ideas.[37] As ideas are tested and pressure is exerted, some are able to withstand the scrutiny whilst others have their weaknesses and flaws exposed and are subsequently defeated. The argument is that the best and most rational idea emerges victorious.

Under this interpretation a deceptive representation can be introduced into public discussion but it will eventually be exposed.[38] The public discussion acts as a "search engine for truth […]",[39] meaning more formalised regulatory mechanisms may be unnecessary.[40] The logic is that if there is sufficient exposure and scrutinisation, "the good will over time drive out the bad and the true prevail over the false".[41] Free speech will provide its own remedy[42] to deceptive speech. Accordingly, restraining or deterring a type of speech is not the solution. Instead, the public debate should be enriched to help identify it. For instance, more speech and counter speech ("communication that seeks to counteract

---

[34] J. S. Mill, *On Liberty* (4th edn, Longmans, Green, Reader and Dyer 1869) pp.94-95.
[35] *Abrams et al v United States* (1919) 250 US 616 Supreme Court no 316, at 630-631 per Justices Holmes and Brandeis.
[36] A. Meiklejohn, *Free Speech and its Relation to Self-Government* (HarperCollins Publishing 2001) pp.26-27. See also A. Meiklejohn, *Political Freedom: The Constitutional Powers of the People* (Harper 1960) p.27.
[37] *Regina (Animal Defenders International) v Secretary of State for Culture, Media and Sport* [2008] UKHL 15, (2008) 1 AC 1312 [28] per Lord Bingham.
[38] *Abrams* at 630-631.
[39] A. Lewis, *Freedom for the Thought That We Hate: A Biography of the First Amendment* (Basic Books 2010) p.185.
[40] A. Meiklejohn, *Free Speech and its Relation to Self-Government* (HarperCollins Publishing 2001) pp.26-27.
[41] *Regina (Animal Defenders International) v Secretary of State for Culture, Media and Sport* n 98 above, [28] per Lord Bingham.
[42] J. Rowbottom, "Lies, Manipulation and Elections— Controlling False Campaign Statements" (2012) 32(3) O.J.L.S. 507, 522.

potential harm that is brought about by other speech")[43] should be added to aid public discussion. As put by Justice Brandeis, "[i]f there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence".[44]

Whilst the self-correcting speech model has appeal, I am not convinced that it is the answer for the same reason other media law scholars have noted e.g., Coe,[45] and Rowbottom.[46] This model is fraught with difficulties and is an idealistic representation of how public debate works. The core issue is that our cognitive processes are not framed in a way which makes us easily-susceptible to changing our beliefs. Individuals are not only biased towards evidence which is familiar[47] but that which fosters confirmation of their previous beliefs.[48] For the most part, even when confronted with evidence to the contrary, the original deceit exerts a "lingering influence on people's reasoning after it has been corrected […]".[49] At times, it can even go further and reinforce the false belief.[50] Afterall, we are particularly predisposed to protect beliefs which form part of our identify (such as political and cultural beliefs). When faced with contradicting evidence[51] we persist in defending our beliefs[52] and will spread the belief further to our "network of like-minded people who have the same propensity to believe […]"[53] to gain affirmation.

---

[43] B. Cepollaro, M. Lepoutre, R. M. Simpson 'Counterspeech' [2023] Philos. Compass. 1, 2.
[44] *WHITNEY v. CALIFORNIA. No. 3*. 274 US 357(1927), 377 per Justices Brandeis and Holmes.
[45] P. Coe, "Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory" (2023) 15(2) J.M.L. 213, 223-227.
[46] Rowbottom, "Lies, Manipulation and Elections— Controlling False Campaign Statements" (2012) 32(3) O.J.L.S. 507, 522-524. See also, I. Katsirea, "Fake news": reconsidering the value of untruthful expression in the face of regulatory uncertainty' (2018) 10(2) J.M.L. 159, 184-185.
[47] U K. H. Ecker, S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga and M. A. Amazeen, "The psychological drivers of misinformation belief and its resistance to correction" (2022) Nat. Rev. 13, 15-16.
[48] Ecker, Lewandowsky, Cook, Schmid, Fazio, Brashier, Kendeou, Vraga and Amazeen, "The psychological drivers of misinformation belief and its resistance to correction" 13. See Rowbottom, "Lies, Manipulation and Elections— Controlling False Campaign Statements" (2012) 32(3) O.J.L.S. 507, 522-523.
[49] Ecker, Lewandowsky, Cook, Schmid, Fazio, Brashier, Kendeou, Vraga and Amazeen, The psychological drivers of misinformation belief and its resistance to correction" (2022) Nature Reviews Psychology 13, 13.
[50] E. Thorson, "Belief echoes: The persistent effects of corrected misinformation" (2016) 33(3) Political. Commun. 460, 461.
[51] D. M. Kahan, 'Misinformation and Identity-Protective Cognition', (Yale Law & Economics Research Paper No. 587, 2017), 5-6 *https://ssrn.com/abstract=3046603*.
[52] Thorson, "Belief echoes: The persistent effects of corrected misinformation" (2016) 33(3) Political. Commun. 460, 461.
[53] Bernal, *The Internet, Warts and All* p.240.

Indeed, the evidence does not support the idea that we have the cognitive processing required for the self-correcting model to work. Numerous political studies, such as those conducted by Nhyan et al,[54] Tabler et al,[55] and Lord et al[56] share the finding that the public is resistant to information which contradicts their beliefs. In these studies, individuals were presented with information which challenged their initial assumptions on controversial issues e.g., capital punishment,[57] weapons of mass destruction, or tax cuts.[58] Regardless of the topic in question, individuals were not convinced by the contradicting evidence. They were resistant to it or worse the contradicting evidence actually reaffirmed their commitment to the belief.[59] In the words of Mackenzie and Bhatt, people are drawn to polarising and false speech like conspiracy theories and generalisations: they are "allured by patently false [information] […] dogmatically persist[ing] with [the] belief […]" in the face of evidence exposing it as false.[60] This, I suggest, demonstrates the general issue with the non-interventionist, self-correcting model. If the public is not able to be open to identifying and correcting their misperception, then the deceit will not be defeated and removed from discussion.

Of course, one may counter this assessment of the public discussion with the admission that yes, the theory in its original form is flawed, but a modified version could be effective. To a degree I can accept this. Studies by those such as Kuklinski et al have demonstrated that the public can be induced to correct their belief.[61] However, the counter speech needs to be more sophisticated than the mere

---

[54] B. Nyhan, and J. Reifler, "When corrections fail: The persistence of political misperceptions" (2010) 32(3) Polit. Behav. 303.

[55] C. Taber and M. Lody, "Motivated Skepticism in the Evaluation of Political Beliefs" (2006) 50(3) A.J.P.S 755.

[56] C. Lord, L. Ross, and M. Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence" (1979) 37(11) J. Pers. Soc. Psychol. 2098.

[57] Lord, Ross, and Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence" (1979) 37(11) J. Pers. Soc. Psychol. 2098, 2105-2107.

[58] Nyhan and Reifler, "When corrections fail: The persistence of political misperceptions" 32(3) Polit. Behav. 303, 314-320.

[59] Nyhan and Reifler "When corrections fail: The persistence of political misperceptions" 32(3) Polit. Behav. 303, 324-320. See also Lord, Ross and Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence" (1979) 37(11) J. Pers. Soc. Psychol. 2098, 2099-2100.

[60] MacKenzie and Bhatt "Bad Faith, Bad Politics, Bad Consequences: The Epistemic Harms of Online Deceit" p.11.

[61] J. Kuklinski, P. Quirk, J. Jerit, D. Schwieder and R. Rich, "Misinformation and the Currency of Democratic Citizenship" (2000) 62(3) J.Politics 790, 792. See also M. Gilens, 'Political Ignorance and Collaborative Policy Preferences' (2001) 95(2) A.P.S.R. 379, 392.

counteraction of a particular view by ordinary citizens.[62] Instead, there needs to be some sort of authoritative statement, provided by an omniscient source, such as state actors acting on society's behalf.[63]

Such an idea, does have a degree of merit. We already have certain organisations in place such as Radio 4's More or Less podcast,[64] and FullFact,[65] which check claims and the use of figures. We even have the Advertising Standards Authority,[66] which has been posed as being appropriate for checking and regulating political advertising.[67] Despite these, we do not have a single authoritative, unbiased and independent body to check political statements. On the odd occasion, we have tried to do so by using the UK Statistics Authority to correct misperceptions. In the midst of the Brexit campaigning, for example, the UK Statistics Authority (an impartial body) intervened and highlighted the limitations of the claim that £350 million will be invested in the NHS if we leave the EU.[68]

However, this example is a good illustration of how even with this limited intervention the narrative can remain uncorrected. In Clarke et al's study, for instance, panellists found that it was "increasingly hard to decipher and believe" the different claims put forward.[69] Part of the issue was that there was no distinguished authority because public figures were throwing their weight behind a wide range of claims. Thus, the clarity which the UK Statistics Authority sought to provide was lost in the midst of discussion. Nevertheless, even if such a body was able to distinguish itself as being *the* leading authority on exposing deceitful representations, it would likely encounter a number of logistical issues if it was used with greater frequency and on a wider scale. Effective countering would require a

---

[62] Rowbottom, "Lies, Manipulation and Elections— Controlling False Campaign Statements" (2012) 32(3) O.J.L.S. 507, 523.
[63] Cepollaro, Lepoutre, Simpson, "Counterspeech" [2023] Philos. Compass. 1, 3.
[64] *https://www.bbc.co.uk/programmes/b006qshd*
[65]*https://fullfact.org/?gad_source=1&gclid=EAIaIQobChMIouPn4Yq6hAMVkphQBh3HnwBDEAAYASAAEgKmGPD_BwE*
[66] Whilst there is a code of best practice, the ASA does not regulate this material, see, ASA Rules 7.1-7.2 at *https://www.asa.org.uk/codes-and-rulings/advertising-codes/non-broadcast-code.html*
[67] This idea was eventually dropped: partly because the ASA primarily deals with commercial advertising (and was ill-suited to regulate non-commercial advertising) and partly because it is financed by advertisers, meaning its independence is more disputable, compared with a statutory body. See, ASA *https://www.asa.org.uk/news/why-we-don-t-regulate-political-ads.html*
[68] UK Statistics Authority, "UK Statistics Authority statement on the use of official statistics on contributions to the European Union"
[69] Clarke, Jennings, Moss and Stoker, "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 114.

burdensome amount of time and resources. This would be both in terms of interrogating the claims made, and then spreading the exposure of the deceit to the public with sufficient accessibility,[70] "reach[,] and frequency".[71] This would take a significant amount of money, time and effort, and even then, some members of the public may not be convinced. Additionally, there may be particular difficulty in correcting the discussion when the deceit is circulated right before a fixed democratic procedure (e.g., days before a referendum or election). In such circumstances, there simply would not be time to investigate and then correct the narrative: the damage would already be done.[72] All this considered, whilst a modified theory of self-correction has appeal, it is not the answer to addressing deceitful political representations. With this analysis undertaken, this brings me to the next section: examining the regulatory mechanisms.

The regulatory mechanisms

The starting point for an analysis of the regulatory part of the framework, are the Principles of Public Life, which unlike measures such as Ministerial Responsibility apply to all public officials. Following the recommendations of the 1995 Nolan report, seven principles were introduced (selflessness, integrity, objectivity, accountability, openness, honesty, leadership)[73] becoming the quintessential basis for what political behaviour should be. A select few have particular applicability to addressing political deceit. The most obvious one is honesty- that public office holders "should be truthful"[74] in the statements they make or promote.[75] There are also associated values of accountability and openness, which encourage transparency ("unless there are clear and lawful reasons for [not] doing so")[76] and the scrutinization of behaviour more broadly. These principles form part of many codes of

---

[70] Clarke, Jennings, Moss and Stoker, "Voter Decision-Making in a Context of Low Political Trust: The 2016 UK EU Membership Referendum" (2023) 71(1) Political Studies 106, 114.
[71] B. G. Southwell, E. A. Thorson and L. Sheble 'Introduction: Misinformation among Mass Audiences as a Focus for Inquiry' in *Misinformation and Mass Audiences* (University of Texas Press 2018) p.5.
[72] Renwick and Palese, "Doing Democracy Better: How Can Information and Discourse in Election and Referendum Campaigns in the UK be Improved?" (The Constitution Unit 2019) p.39.
[73] The Committee on Standards in Public Life, *Standards in Public Life Vol 1,* p.14 para 55.
[74] GOV.UK, "The Seven Principles of Public Life" (31 May 1995) *GOV.UK*, *https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2*
[75] R. Thomas, 'Fake news and the Nolan Principles' *Committee on Standards in Public Life* (6 March 2017) *CSPL, https://cspl.blog.gov.uk/2017/03/06/fake-news-and-the-nolan-principles/*
[76] GOV.UK, "The Seven Principles of Public Life"

conduct, including the Ministerial Code, House of Commons and House of Lords, which all hold an overarching duty to observe them.[77]

For the most part, non-compliance with these principles does not result in punitive action. The principles tend to be used as part of an integrity-based model,[78] relying on individual consciences, rather than an overseeing body to regulate behaviour. The premise is that ethical principles are not enforced but embedded into political culture,[79] over time becoming internalised norms of what is acceptable.[80] However since their introduction they have not prompted a drastic change in behaviour,[81] mainly because they rely on individual consciences to uphold them.[82] As MP Liz Saville puts, "we are no longer in [a] […] world [where] chivalry and words as bonds […]"[83] are enough to regulate most political behaviour.

Although the Principles of Public Life mainly operate as guidance for political behaviour, they do have some capacity to trigger sanctions. Of note, is the Ministerial Code where these are enforceable as standards of behaviour. There is not only an overarching duty on Ministers "to adhere to the Seven Principles of Public Life […]",[84] but a specific stipulation that Ministers should be as open as possible with Parliament and the public. Refusal to provide information is only permissible when disclosure would not be in the public's interest.[85] Breaching these Ministerial standards can result in an

---

[77] Cabinet Office, "Ministerial Code" (November 2024) para 1.4 *Cabinet Office*, *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1126632/Ministerial_Code.pdf.* See also House of Lords, *Code of Conduct for Members of the House of Lords, Guide to the Code of Conduct, Code of Conduct for House of Lords Members' Staff* (2022, H.L. 13) pp.2-3 paras 9-12, p.10 para 11; House of Commons, *The Code of Conduct together with The Guide to the Rules relating to the Conduct of Members* (H.C. 2023, 1083) pp.2-3.
[78] F. M. Lartey "INTEGRITY-BASED AND COMPLIANCE-BASED ETHICS PROGRAMS: A CRITICAL ANALYSIS OF KEY DIFFERENCES" (2021) 5(5) I.J.B.E.M. 43, 44
[79] The Committee on Standards in Public Life, *Standards in Public Life Vol 1,* p.3 para 6.
[80] P. Spicker, "Seven Principles of Public Life: time to rethink" (2014) 34(1) P.M.M 11, 11-12.
[81] M. Bull "Whatever happened to the Nolan principles? Sleaze in the government of Boris Johnson" (17 May 2021) *LSE*, *https://blogs.lse.ac.uk/politicsandpolicy/nolan-sleaze-johnson/*
[82] R. Roberts, "The rise of compliance-based ethics management" (2009) 11(3) Public Integrity 261, 261-273.
[83] Elected Representatives (Prohibition of Deception) Deb Tuesday 28 June 2022, cols 183-185 (Liz Saville).
[84] Cabinet Office, "Ministerial Code" para 1.4. See also Independent Advisor on Minister's Interests, 'Independent Adviser on Ministers' Interests Annual Report 2022-2023' (May 2023) Independent Advisor on Ministers' Interests, para 1 *https://www.gov.uk/government/publications/annual-report-of-the-independent-adviser-on-ministers-interests-may-2023/independent-adviser-on-ministers-interests-annual-report-2022-2023-html*
[85] Cabinet Office, "Ministerial Code", para 1.4(e).

investigation and sanction (e.g., a public apology, remedial action, removal of ministerial salary, or removal from office). Although the Prime Minister can ask the Independent Advisor to investigate the matter and the Independent Advisor can make recommendations, the ultimate judge of whether there has been a breach and what the consequence should be is the Prime Minister.[86] One would think that based on this, that this mechanism would be effective at addressing deceit made to the public. Afterall, if the public is deceived by a Minister, then this is a potential breach.

However, this is not how it works in practice because the Prime Minister does not enforce this standard. Under this mechanism, the Prime Minister is the judge of what is acceptable, and tends not to open investigations just because a Minister deceives the public. Even if the Prime Minister did decide to do so, they are able to opt for minimal sanctions or ignore a finding that a breach occurred.[87] Although the Prime Minister usually follows the advice of the Independent Advisor, there are instances where the findings of their investigation have not been accepted. One notorious example, is Boris Johnson who ignored his then-advisor's findings that Priti Patel was bullying members of staff, allowing her to breach the code without repercussion.[88]

Some of the problem may be ameliorated with the recently updated Ministerial Code. Now, the Independent Adviser greater autonomy and the power to launch investigations into potential breaches. This means that instances of deception which had previously not been investigated may now gain attention and be looked into. Yet, this really just bolsters the investigative side of the issue. The actual enforcement side remains deficient and still grants a great deal of power to the Prime Minister. Decisions over whether a breach has occurred, whether to impose a sanction and the form it should take all lie with the Prime Minister.[89] Regardless of the nature of the problem, the point I am making

---

[86] Cabinet Office, "Ministerial Code" para 2.7.
[87] S. Blewett, D. Lynch, D. McGrath, "Rishi Sunak appoints ethics adviser but accused of preserving 'rotten regime'' (22 December 2022) *Press Association* *https://www.proquest.com/docview/2756691907?parentSessionId=aS1DldJyXi6ckxmlRJXui4mKevTYJidbaoEyYnh539E%3D&pq-origsite=primo&accountid=14533*
[88] S. Murphy "Alex Allan: the veteran windsurfing mandarin who quit over Patel row" (20 November 2020) *The Guardian*, *https://www.theguardian.com/politics/2020/nov/20/alex-allan-the-veteran-windsurfing-mandarin-who-quit-over-bullying*
[89] Cabinet Office, "Ministerial Code" paras 2.6-2.7.

is that the Ministerial Code and the Principles of Public Life are not appropriate for addressing deceitful representations made to the public. The insufficiency of the existing regulatory mechanisms is one that extends beyond the Ministerial Code and continues into the legal remedy (section 106 of the Representation of the People Act 1983).

Section 106, is integral for any analysis of regulating deceitful representations made to the public, because it is the only legal mechanism which specifically combats it (albeit only electoral defamation). To quote section 106, this provision makes it an illegal practice for a person to make or publish "any false statement of fact in relation to [a] candidate's personal character or conduct […]",[90] "before or during an election",[91] "for the purpose of affecting the return of any candidate at the election".[92] Unless, the individual can show that they "had reasonable grounds for believing, and did believe, the statement to be true".[93] It draws together criminal penalties such as prosecution for an illegal practice and on a summary conviction a fine not exceeding level 5,[94] with civil remedies like the claimant can apply for an injunction to restrain distribution of the deceit.[95] In addition, electoral sanctions can be imposed such as the election being declared void[96] and/ or the respondent being barred from standing for Parliament (or holding elective office) for three years, as well as being asked to vacate the seat.[97]

It is easy to note similar issues to those which were identified in relation to the Principles of Public Life and the Ministerial Code. For instance, it is evident that section 106 has limited applicability, and only addresses a certain variety of deceitful representation made by politicians to the public. Only targeting electoral defamation is itself enough to warrant criticism as it narrows the scope of what is being addressed substantially, but the reality is the scope is much more limited. One reason is that the

---

[90] Representation of the People Act 1983, s.106(1)(b).
[91] RPA 1983, s.106(1)(a).
[92] RPA 1983, s.106(1)(b).
[93] RPA 1983, s.(1)(b).
[94] RPA 1983, s.169. See also *DPP v Edwards* [2002] EWHC 636, [2002] 3 WLUK 551 at [4]-[13].
[95] RPA 1983, s.106(3).
[96] RPA 1983, s.135, s159.
[97] RPA 1983, s.106(2), s.160(4), (5)(b), s.173(1)-(3). See also Rowbottom, "Lies, Manipulation and Elections—Controlling False Campaign Statements" (2012) 32(3) O.J.L.S. 507, 508.

illegal practice only applies to the personal conduct of a fellow candidate.[98] The courts have repeatedly emphasised that there is a distinction between "whether the statement is one as to personal character or conduct[,] or a statement as to the political position or character of the candidate".[99] The former is actionable and the latter is not. The second limiting factor is that section 106 is only applicable in cases where there has been a lie. As argued by Horder, the notion of "a false statement of fact in relation to the candidate's personal character or conduct" notably excludes deception which is misleading. "[S]omeone might disseminate a perfectly correct claim that a candidate has been convicted of murder, whilst failing to mention that the conviction was later overturned and the candidate completely exonerated [...]".[100]

The narrowness of this scope has meant that section 106 has had little effect. Data from both the Electoral Commission and the law reports prove that this offence has had little use as a legal remedy. For instance, there have only been two convictions in the last thirteen years (despite over 800 allegations having been made to the police during this time).[101] Similarly, only a handful of electoral petitions have been brought.[102] Now, part of this inefficacy may be addressed if the scope of the offence was widened e.g., to accommodate political defamation and other types of representations (like misleading statements). However, even if these amendments were made what it would be addressing, is still negligible. The core issue is, that this mechanism would still only be addressing electoral defamation and not representations which are made to the public more broadly. Much like the Principles of Public Life and the Ministerial Code, even with modifications section 106 would not go far enough.

---

[98] RPA 1983, s.106(1)(b). See also *Cumberland Cockermouth Division case* (1901) 5 O'M&H 155, 160.
[99] *Regina (Woolas) v Parliamentary Election Court* [2010] EWHC 3169, (2011) 2 W.L.R. 1362 at [111]. See also [107]-[111], [114].
[100] J. Horder, *Criminal Fraud and Election Disinformation: Law and Politics* (Oxford University Press 2022) p.146.
[101] Electoral Commission, "Fraud data 2010-2016" (Electoral Commission, 2023); Electoral Commission, 'Electoral Fraud Data 2017-2023' (2024) *Electoral Commission*, *https://www.electoralcommission.org.uk/search?search=fraud+data*
[102] Only 8 civil actions have been brought on s106 grounds in the last thirteen years (electoral petitions or applications for interim injunctions) and few have been successful. Based on a search of the Law Reports through the online databases of Westlaw last conducted in November 2024. Note, that I excluded appeals.

My overarching point in this part of the paper is that the current approach is failing: none of the mechanisms are effectively addressing deceitful representations. Whether it be a lack of comprehensiveness or a lack of use there is inadequate protection in place. Based on this analysis, I suggest that reform is needed. Specifically, I advocate for a new specific criminal offence to complement the existing framework. This would help to ensure that the public's formation of political preferences and democratic participation is based on truthful information.

Towards a Criminal Offence

Having carefully assessed the framework and identified its deficiencies, I will now turn to setting out how we should move forward, specifically suggesting a new criminal offence. I begin by setting out a rough conceptualisation of what I envision the offence to be and then move onto offering some analysis on it, setting out why it is appealing.

A new specific offence for the most egregious representations

In this paper, my argument is that we should introduce a new specific criminal offence for some of the deceptive representations which are made to the public by politicians. Whilst it is beyond what can be achieved in this paper to offer a draft offence, it is necessary to offer a rough indication of what I envision the offence to be. Although I readily admit that this is not a complete outline, I do seek to offer an indication of what the offence should be capturing and highlight certain principles which should form its foundation.

In my view the remit of the criminal offence should be limited to the most egregious representations similar to instances like Boris Johnson claiming that "we send the EU £350 million a week",[103] or Tony Blair claiming that "we have the intelligence that says that Saddam has continued to develop these weapons of mass destruction […]".[104] In my view these are the most egregious because they embody a number of principles which are indicative of greater culpability and potential for significant harm.

---

[103] Asthana, "Boris Johnson: we will still claw back £350m a week after Brexit".
[104] BBC, "Breakfast with Frost".

First and foremost, egregious instances are forms of true deceit- the politician is aware that what they are saying was false or misleading, or knew that there was risk of it being so. They are therefore worthy of blame because there is a duplicity between what they know to be true and what they are inducing others to believe.

Second, such representations have the potential to cause the most harm to the formation of political opinions and democratic engagement. This requires the following conditions. One, the representation needs to be made by a certain type of politician, namely they need to hold parliamentary or governmental office. Such roles tend to carry greater political influence and are consequently more likely to be given greater credibility by the public when they form their political preferences.
Two, the representation needs to have the potential for collectivised harm. Thus, it needs to be made to the public more broadly, or the politician needs to be aware that it will likely be disseminated to the public more broadly in the near-future. So, this could include conventional communicative transactions e.g., speeches, social media or press releases, but it could also involve other means e.g., radio or television interviews. The obvious exclusion to the behaviour would be speech which is covered by parliamentary privilege (on account of the need for free parliamentary debate).

Third, the representation needs to be a material matter of public interest- as in relevant to the public and with the potential to be influential on their decision-making.

Aside from basing the offence around principles which capture the greater propensity for culpability and harm, I would also include an excuse for justification for justifiable deceit to reflect some situations where the deceitful representation was required or necessitated. The classic example to this would be in instances of national security or when secret negotiations are taking place.
Ultimately, this is what I suggest makes deceptive representations so egregious that they warrant a response. It is these very principles which I suggest should form the basis of the offence.

A case for this new offence

In my view it is natural to turn towards using the criminal law to address this issue, as opposed to other regulatory mechanisms e.g., parliamentary committees or ordinary regulators. Compared to the

other options, criminal law is more expressive and symbolic- it is after all the most severe expression of societal disapproval. The criminal law achieves this through a systematic and holistic reflection on the behaviour, beginning with the stigmatisation associated with the act. The act of criminalisation is a declaration that the state sees it as wrongful and society recognises the severity of being labelled a criminal (e.g., having a criminal record). Being labelled a criminal is something which needs to be disclosed in certain formal contexts (e.g., employment and visa applications).[105] Of course, the punishment itself then also acts as a severe form of censure.[106] Whilst other political and legal mechanisms do have sanctions attached (e.g., typical political sanctions involve a formal reprimand, suspension, or loss of pay and civil remedies involve damages) they are generally not as costly as criminal sanctions. My point is that the criminal law has a social significance and resonance[107] which other regulatory mechanisms do not carry. It sends a powerful message that the conduct will not be tolerated and that if someone engages in this behaviour then they will be labelled a criminal.[108]

This expressive quality should mean that the criminal law is effective as a deterrence. What I mean by this, is that the cost of being labelled a criminal and the actual sanctions imposed should mean that it has a significant cost thereby having a great deterrence on behaviour. The deterrence model is based on rational cost-benefit analysis e.g., if you increase the cost of something, less of it will be consumed. So, if you make a behaviour more costly to engage in less of it will occur[109] because the benefit of engaging in the behaviour is outweighed by the burdens of potentially being caught and punished.[110] If we draw on this reasoning then it is natural to turn towards the more costly mechanism of a criminal offence for political deceit, compared to say regulation through a parliamentary committee or an ordinary regulator. In turn, this should reduce the amount of political deceit that occurs. The main attraction of using this model, is that it works to prevent the behaviour from

[105] A. P. Simester and A. von Hirsch, *Crimes, Harms and Wrongs: On the Principles of Criminalisation* (Hart Publishing 2011) p.4.
[106] Simester and von Hirsch, *Crimes, Harms and Wrongs: On the Principles of Criminalisation* p.5.
[107] Simester and von Hirsch, *Crimes, Harms and Wrongs: On the Principles of Criminalisation* pp.4, 212.
[108] C. McGlynn, "Challenging anti-carceral feminism: criminalisation, justice and continuum thinking" Women's Studies International Forum (2022) 93 102614.
[109] G. Tullock, "Does Punishment Deter Crime?" (1974) 36 The Public Interest 103, 105.
[110] T. Brooks, *Punishment: A Critical Introduction* (2nd edn, Taylor & Francis Group 2019) p.45.

occurring in the first place, and as a result is forward-looking.[111] The fact that people are often unwilling to change their mind once they have formed a political opinion means that it is logical to try and pre-empt it: preventing the behaviour from happening and the harm before it occurs. Of course, we also need to impose consequences which reflect the gravity of one's political deceit[112] (retribution) but the strategy of the regulation should centre around deterrence, as opposed to gearing towards a backward-looking response, particularly in light of the idea of protecting the public's ability to form political preferences and engagement in democratic procedures.

Now one may counter this with the argument that deterrence theory is not that successful, drawing on empirical studies to do so. For example, Anderson's survey into male prisoners and the factors influencing their offending suggests that most criminals do not contemplate the potential effect of the criminal behaviour before engaging in it. He notes that "76% of active criminals and 89% of the most violent criminals either perceive no risk of apprehension or are incognizant of the likely punishments for their crimes".[113] With that being said, there is evidence to suggest that the success of deterrence theory varies by context, with a propensity for it to be more effective in cases of minor or administrative crime when individuals undertake rational choice theory before the act i.e. weighing the risks up against the benefits. Dölling et al's meta-analysis of 700 studies [114] notes the effect of the deterrence model varies by type of crime. They note that "statistically significant estimations [are] […] to be found [in minor crimes like] […] traffic offences, whereas the deterrence hypothesis is rarely confirmed in the case of more serious offences". They attribute this to the role of rational choice, which is more prevalent in minor offences, whereas major crimes are often characterised by expression e.g., emotion and spontaneity which often do not give the opportunity for rational thinking

---

[111] Brooks, *Punishment: A Critical Introduction* p.44.
[112] Brooks, *Punishment: A Critical Introduction* pp.48-49.
[113] D. A. Anderson, "The Deterrence Hypothesis and Picking Pockets at the Pickpocket's Hanging" (2002) 4(2) American Law and Economics Review 295, 295.
[114] D. Dölling, H. Entorf, D. Hermann and T. Rupp, "Is Deterrence Effective? Results of a Meta-Analysis of Punishment" (2009) 15 Eur J Crim Policy Res 201, 214-215.

and a risk assessment.[115] Similar findings can be seen in Abramovaite et al's[116] analysis of police forces and the reduction of theft, burglary and violence in England and Wales. Yes, increasing the certainty of punishment (measured by increased police detections) was associated with a reduction in non-expressive crimes like theft and burglary, but it did not have the same effect on violent crime.

Now, deceptive representations are not crimes characterised by emotion or spontaneity so they are an act which would naturally lend themselves towards a risk assessment. Further, once the politician engages in the risk assessment there are a number of factors that would colour their perception. First, a politician is likely aware of the risks and penalties attached to the offence e.g., likely punishment, so will be well-informed of the potential costs and fully understand the implications of getting caught.[117] Second, politicians have a high social and economic status which should mean that they have more to lose if they engage in criminal acts.[118] Those "who receive relatively few rewards from society, whether [it be] economic or social, would tend to place a greater value on the potential rewards for criminal activity". Conversely, those who receive more rewards will perceive "greater informal costs […]"[119] from engaging in the behaviour such as loss of status and prestige. The political profession aligns with the latter and will mean that they potentially have a lot to lose from engaging in criminal behaviour. In particular, the publicity that is associated with the role, and the long-term effect on their career and livelihood that being caught may have[120] are notable costs and should help to act as a significant deterrent.

---

[115] D. Dölling, H. Entorf, D. Hermann and T. Rupp, "Is Deterrence Effective? Results of a Meta-Analysis of Punishment" (2009) 15 Eur J Crim Policy Res 201, 215. See also D. A. Anderson, "The Deterrence Hypothesis and Picking Pockets at the Pickpocket's Hanging" (2002) 4(2) American Law and Economics Review 295, 298, 301-304.

[116] J. Abramovaite, S. Bandyopadhyay, S. Bhattacharya, and N. Cowen, "Classical deterrence theory revisited: An empirical analysis of Police Force Areas in England and Wales" (2023) 20(5) European Journal of Criminology 1663, 1663.

[117] P. H. Robinson and J. M. Darley, "Does Criminal Law Deter? A Behavioural Science Investigation" (2004) 24(2) O.J.L.S.173, 176.

[118] M. R. Geerken and W. R. Gove, "Deterrence: Some Theoretical Considerations" (1975) 9 Law Society Review 497, 498, 507.

[119] Geerken and Gove, "Deterrence: Some Theoretical Considerations" (1975) 9 Law Society Review 497, 507-508.

[120] Geerken and Gove, "Deterrence: Some Theoretical Considerations" (1975) 9 Law Society Review 497, 503-507.

I do admit that the deterrence-effect may be reduced once a person of high socio-economic status is convicted of the offence. As is evident in Weisburd et al's 1995 study into white collar criminals, harsher sanctions like imprisonment had no major effect on reoffending[121] however, this is likely because the costs of being caught become less significant once there has been that first conviction. Their reputation or employment prospects will have already been damaged. My point is not so much that people who have already been convicted will be deterred from engaging in this type of speech, but that those who are not yet offenders will be deterred because the stakes are still high. For those who are not already offenders, the deterrence effect should work.

The overarching point here, is that I have a broad preference towards criminalisation- the criminal law should be more effective at reducing the number of politicians who make deceitful representations, particularly when compared with other regulatory mechanisms. With that being said, there are a number of objections which could be raised in response to my suggestion of using the criminal law. I will examine and address the most frequently raised objections, seeking to highlight the appeal of a new specific offence.

The first objection which could be levied at criminalisation is what I term the legislative objection. What I mean by this, is that it would require the very people who the Act would be restricting to support its passage. I can admit that typically, Parliament has been unwilling to support the passage of similar legislation. For instance, there have been attempts to pass Bills making it illegal for politicians to make deceptive statements (the Elected Representative (Prohibition of Deception) Bill was introduced in 2006 and 2022 and both times the Bills failed to get a second reading). Whilst I admit that there will be a degree of self-interest to overcome, there is reason to believe that Parliament could support a new and different offence. The offence I set out is significantly narrower compared to the one in either of the Elected Representative Bills. Even in the rough outline I provided there are extra conditions of materiality and a broader justifiability defence. As the offence will have less general applicability, it should be more appealing to those in Parliament.

---

[121] D. Weisburd, E. Waring and E. Chayet, "SPECIFIC DETERRENCE IN A SAMPLE OF OFFENDERS CONVICTED OF WHITE COLLAR CRIMES" (1995) 33 Criminology 587, 587, 593-597.

With that being said, there is evidence to suggest that Parliament can set aside its self-interest. What is particularly indicative is the recent move by the Welsh government committing to making lying in politics illegal.[122] Although what this will entail is unknown (the legislation has not been drafted) the fact that politicians are willing to commit to it and support its passage, suggests that Parliament introducing an Act to criminalise this behaviour is possible. Setting aside these more specific reasons, it is important to stress that most British politicians are answerable to the public. If the public feels that their current representatives are not serving their interests, then politicians can lose political power at the next election. Therefore, if enough public pressure is exerted and expressed in favour of a Bill to address this issue, then politicians would be forced to support it because a failure to do so would risk their political power.

Another possible objection is what I term the politicisation objection. As in the argument that creating a criminal offence to address political deceit would encourage the judiciary to become overly involved in political matters when these institutions should be kept separate. Whilst I can appreciate the validity of this objection, what I am advocating for is a very narrow offence. It would only involve expanding the judiciary's role in addressing political behaviour to a very small degree. Indeed, we already have a number of criminal provisions which each regulate very small amounts of behaviour from our political representatives and public figures. Again, I refer to section 106 of the Representation of the People Act 1983 as well as other offences such as bribery[123] and various provisions which relate to election expenses.[124] My point is that the law already addresses other political issues and has not been brought into disrepute. If you consider this, in light of the fact that the offence would be very narrow and not be something which the courts would encounter every day, then the judicial role would not be extended very much and would likely not influence public perception on judicial independence.

[122] S. Morriss, "Welsh government commits to making lying in politics illegal" (2 July 2024) *The Guardian*, *https://www.theguardian.com/politics/article/2024/jul/02/welsh-government-commits-to-making-lying-in-politics*
[123] Bribery Act 2010, s.1-2, 6. See also RPA 1983, s.113 for bribery in the context of voting.
[124] RPA 1983, s.72-90D.

A further objection which could be made is what I term the logistical objection. As in the question could be posed how much would difference would such a narrowly drafted offence make? I concede that the offence is quite narrow and this would be compounded by the fact that it is a difficult process to actually enforce the offence. For instance, the Crown Prosecution Service (CPS) requires there to be a realistic possibility of conviction and it to be in the public interest before proceeding with prosecution. Further, even if they do decide to progress with prosecution, court proceedings can take a long time by which point the damage is done. I can appreciate both these objections and do acknowledge that if imposed, this offence would have limited applicability and be subject to the logistical limitations of legal proceedings. With that being said, I am not proposing something that would be used frequently, instead, I am suggesting something which deliberately has limited applicability. Its purpose is not to address deceptive representations more broadly but is more narrowly construed so as to only address the most egregious representations and the most clear-cut cases. So, in my view it is right that this offence would have limited use. In regards to the second issue of it potentially taking a long time to proceed to court, I can fully admit that this is a valid point. However, this seems to be a natural limitation of using any legal mechanism to address behaviour, it does not mean that we should not use them.

The final objection which could be levied at criminalising certain types of political deceit is the free speech critique. What I mean by this, is that you could argue that criminalising a type of political speech would be at odds with our international obligations under the European Convention of Human Rights (ECHR) and in light of the European Court of Human Rights (ECtHR) jurisprudence. To a degree, I can understand the concern, but I do not agree with the sentiment that criminalising this type of political speech would be a violation.

Under Article 10 of the ECHR[125] there is a general guarantee afforded to free expression and political expression is guarded even more fervently with less capacity for interference something which is reflected in ECtHR case law. Traditionally, the court has been very reluctant to grant interferences,

---

[125] European Convention on Human Rights 1950, Article 10.

offering protection to an entire spectrum of political speech from information or ideas which are favourably received to those which are indifferent and even to those which offend, shock or disturb.[126] As such, not only have relatively straightforward types of political speech been protected (such as ideas which challenge the current institutional order),[127] but also more controversial and offensive types such as where a political applicant engaged in hateful, hostile or offensive rhetoric.[128] More importantly for the purposes of this paper, ECtHR jurisprudence seems to support the fact that Article 10 protection extends to making false political claims. The pertinence of this is clearly illustrated by *Salov v Ukraine.*[129] In this case Mr Salov (the applicant) had disseminated incorrect information about the alleged death of a presidential candidate, thereby breaching Article 127 of Ukraine's Criminal Code by interfering with "electoral rights, or […] with the activity of an electoral commission, for the purpose of influencing election results […]"[130] and as a result received criminal penalties for doing so.[131]

The Court seemed to have no issue with the fact that the state had satisfied the first two limbs of the test for legitimate interference under Article 10(2). First, the Court found that the legislation under Article 127 was sufficiently foreseeable and clear so as to allow an individual to see the consequences of their behaviour.[132] To add to this, Article 127 was a measure which imposed sanctions for speech that was already made and the ECtHR generally tends to be more tolerant of measures restraining speech.[133] Second, the Court affirmed that safeguarding against false or misleading political

[126] *Eğitim ve Bilim Emekçileri Sendikası v. Turkey* App no 20641/05 (ECtHR, 25 September 2012) para 67, citing *Handyside v United Kingdom* App no5493/72 (ECtHR, 7 December 1976) para 49, and *Jersild v Denmark,* App no 15890/89 (ECtHR, 23 September 1994), para 37.
[127] *Eğitim ve Bilim Emekçileri Sendikası v. Turkey*, paras 70-74.
[128] See, for instance, *Jersild v Denmark, GÜNDÜZ v. TURKEY* App no 35071/97 (ECtHR, 4 December 2003); *Erbakan v Turkey* App no 59405/00 (ECtHR, 6 July 2006).
[129] *Salov v Ukraine* App no 65518/01 (ECtHR, 6 December 2005).
[130] Criminal Code of Ukraine, Art 127.
[131] *Salov v Ukraine*, paras 10-32.
[132] *Salov v Ukraine,* paras 108-110.
[133] *Savva Terentyev v Russia*, App no 10692/09 (ECtHR, 28 August 2019) para 65.

information fell under the legitimate aim[134] of protecting the publics engagement with the democratic processes (such as choosing a presidential candidate).[135]

With that being said, the major issue the Court had was that the interference failed on the latter part of the test: being necessary in a democratic society and proportionate to the aim pursued.[136] In relation to necessity, the Court drew particular attention to the nature of the speech in question, noting the fact that the information was not produced or published by the defendant. Instead, it was; a personalised assessment, had limited impact, and lacked evidence of being made with deceitful intention.[137] When these characteristics were collectively considered in light of the fact that Article 10 protects the discussion or dissemination of information (even if it is strongly suspected to be untrue), there was no need for the interference.[138] Aside from the issues with necessity, the measures imposed were seen as disproportionate to the aim pursued. In particular, the nature and severity of the penalties imposed far outweighed the aim of ensuring the democratic process.[139] In this case, Mr Salov was given a sentence of five years (which was suspended for two), a fine, and annulment by the Bar Association of the applicant's licence to practise law.

Now it is important to stress that I agree with the Courts approach in *Salov*: for that particular speech the measure was unnecessary and the sanctions imposed disproportionate.[140] However, this case can be interpreted as suggesting that there may be circumstances in which political deceit *may* in fact be criminalised in a way which is compatible with Article 10(2).[141] The issue of proportionality may be fairly easily resolved by introducing an offence with less severe sanctions e.g., fines or community orders as opposed to imprisonment, but the problem of necessity may require more work.

I fully admit that not all low value speech (such as false or deceitful political rhetoric) poses such a

[134] *Salov v Ukraine,* para 110. See also *AHMED AND OTHERS v THE UNITED KINGDOM* App no 65/1997/849/1056 (ECtHR, 2 September 1998), para 52 where the Court stressed the importance of ensuring the free will of the people during elections and the need to protect democratic society from interferences.
[135] *Salov v Ukraine,* para 101.
[136] *Salov v Ukraine,* para 116.
[137] *Salov v Ukraine,* paras 113-116.
[138] *Salov v Ukraine,* para 113.
[139] *Salov v Ukraine,* para 115.
[140] *Salov v Ukraine,* paras 110-113.
[141] See Coe's different interpretation of the case in Coe, "Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory" 10-11.

significant threat to the democratic process as to necessitate criminalisation. With that being said, the Court's approach may be different if there was a change in the nature of the deceit. Say, for example, the individual was more culpable, or the deception had more potential for harm by embodying certain qualities which exacerbated its threat. Whilst this is something which was not met by the facts in *Salov*, the reasoning underpinning the judgement suggests that there are other factors which may sway their assessment of whether criminalising political deceitful representations unduly interferes with Article 10.[142] The point to be taken is the offence would need to be more discerning with what it addresses compared to *Salov*, something which I suggest my proposed offence achieves. My offence is very narrowly drafted and only seeks to address the instances of deceit which are indicative of higher levels of culpability and harm (the most egregious). In turn, these are more likely to satisfy the requirement of necessity. The point I am making is that even though the ECtHR has previously viewed similar interferences as violating Article 10, there is reason to argue that the opposite can be achieved, particularly if you impose a narrower and carefully drafted offence.

Whilst I willingly accept that criminalisation is not a route to be entered into lightly (and that there are some legitimate concerns), I have argued that there is a case for criminalising the most egregious instances of deceitful representations and highlighted the appeal of such an offence.

Conclusion

Recent shifts in the political sphere have cast deceptive representations into a new light. As time has progressed, greater scrutiny has uncovered the regularity with which deceptive representations are being made. With the extent that politicians are willing to deceive us now being exposed, the democratic implications on the public's political preferences become more acute. Such developments create an opportunity to reflect on the UK's approach and ask whether it is suitable.

A careful assessment of the UK's approach to addressing deceptive representations highlights how the framework is struggling to cope. I suggest that a new response is needed to help complement what is

[142] See similar points being made about a more discerning threshold in the US context, S. Lieffring, "First Amendment and the Right to Lie: Regulating Knowingly False Campaign Speech after United States v. Alvarez Note" (2012-13) 97 Minnesota Law Review, 1047, 1061-1076.

already in place. My suggestion is that the most effective response would be a new criminal offence. I admit that the criminalisation route is something which has been attempted before and often encounters objections. Yet, I still argue that we should use it but take a different tact. Developments in ECtHR jurisprudence demonstrate that a new and narrower criminal offence may be more appropriate. Particularly towards the end of this paper, I put forward a rough idea of what I envision the offence to be. Here, I emphasise how the specificity of the offence seems to mollify some of the frequently cited concerns about criminalising this type of behaviour and stress how it is a feasible course of action. My hope is that following the changes in the political sphere, we reassess our view of the feasibility and capability of an offence to address deceptive representations. I suggest that it has untapped potential in stemming their use.

Document 7 - The Law Society

Written evidence: Senedd Members Standards of Conduct Committee Inquiry into Members' Accountability

05 December 2024

The Law Society

Introduction

This submission provides an initial analysis of the proposals in the Institute for Constitutional Democratic Research (ICDR) White Paper, *"A Model for Political Honesty,"* within the context of the Senedd Members Standards of Conduct Committee Inquiry into Members' Accountability. This response specifically addresses our members thoughts around the potential implications on fairness, legal standards, and practical enforcement.

1. Misalignment with Objectives

The White Paper outlines a legislative framework aiming to disqualify politicians found guilty of deliberate deception. While the stated intention is to penalise deliberate dishonesty, the mechanisms appear to be misaligned with this objective in the following ways:

a) Absence of Clear Deliberate Deception Standard

- The White Paper does not adequately ensure that a finding of deliberate deception, as understood in legal standards, is a prerequisite for disqualification. Instead, it conflates:
    - A failure to correct statements deemed false on a balance of probabilities, and Intentional acts of deception.

b) Undermining Accountability Goals

- By failing to distinguish intent from error or oversight, the proposed model undermines both the clarity of the process and the confidence of the public in its fairness. Political discourse is inherently complex, and allegations must be adjudicated based on evidence of intent, not merely incorrectness.

2. Standards of Proof and Procedural Safeguards

The principles of justice necessitate adherence to appropriate evidentiary standards and procedural safeguards. The ICDR Model fails to address these adequately:

a) Burden of Proof

- Allegations of deliberate deception carry severe reputational and professional consequences. They must therefore meet the **criminal standard of proof**, where the tribunal of fact must be sure beyond a reasonable doubt.

- The proposed reliance on the civil standard of balance of probabilities is inadequate and risks unjust outcomes, particularly where complex evidentiary issues arise.

b) ***Legal Precedents***

- Case law such as *Re B (Children)* [2008] highlights the importance of higher evidentiary standards for serious allegations, even in civil cases. This principle should be mirrored in procedures involving public officials.

- The concept of a "lie," as defined in *Lucas (1981)*, requires intent. By allowing findings of dishonesty without proving such intent, the ICDR Model strays from established legal definitions.

3. Procedural and Resource Challenges

The ICDR Model introduces procedural frameworks that are not only novel but also likely to impose severe burdens on the judicial system, particularly:

a) Impact on Magistrates' Courts

- The White Paper proposes a "leave" stage for District or Deputy District Judges to review applications within 24 hours. This places significant additional strain on already overstretched Magistrates' Courts.

- At the end of 2023, backlogs in Magistrates' Courts in Wales reached 17,480 cases, the highest recorded since 2012, with the situation worsening. Adding a high-volume, expedited process risks overwhelming the system further.

b) Resource Implications for Appeals

- The suggested appeal process to the Crown Court on points of law introduces a layer of complexity and delays, with Crown Court backlogs also at unsustainable levels (2,595 cases in Wales as of December 2023).

- Without significant investment in judicial capacity, these procedural demands are unlikely to achieve the stated aim of swift resolution.

c) Increased Litigation Risks

- The novelty of the ICDR Model, including concepts such as the "real possibility" test, invites legal challenges. The lack of clarity in its framework may lead to prolonged disputes and delays, compounding the resource burdens on the justice system.

4. **Broader Implications**

The ICDR Model, if implemented, risks creating significant unintended consequences:

a) Effect on Political Discourse

Political discourse often involves complex, contested claims. Penalizing statements without clear evidence of deliberate deception risks deterring open debate, an essential feature of democratic systems.

As noted by the Electoral Commission, political campaigns involve presenting data in ways favourable to one's argument. The subjectivity inherent in such presentations must be recognized to avoid criminalizing routine political expression.

b) Existing Legal Protections

- These existing mechanisms balance the need for accountability with fairness, requiring proof of intent and providing safeguards against misuse.

c) Practical Barriers to Implementation

- The White Paper fails to consider the administrative and procedural burdens associated with its proposals, including the risk of vexatious or coordinated complaints overwhelming the system.

- The lack of adequate filtering mechanisms exacerbates these risks, increasing the likelihood of frivolous or unmeritorious cases proceeding unnecessarily.

Potential Next Steps

To achieve the stated goal of enhancing accountability while maintaining fairness and practicality, the following recommendations are proposed:

a) Strengthening Existing Frameworks:

- Refine existing disciplinary procedures to include robust sanctions for deliberate deception without creating parallel legal frameworks that dilute standards of proof.

b) Clear Evidentiary Standards:

- Any new procedures must require proof of intent to deceive, meeting the criminal standard of proof where necessary.

c) Judicial and Resource Considerations:

- Conduct a comprehensive impact assessment to evaluate the resource implications of any proposed changes, ensuring that judicial and administrative capacity can meet the demands.

d) Safeguards Against Misuse:

- Introduce strict thresholds for initiating complaints and mechanisms to filter out frivolous or vexatious claims early in the process.

e) Stakeholder Engagement:

- Collaborate with legal and civic organizations to design practical accountability mechanisms that are both fair and efficient.

Conclusion

The aspiration to improve accountability in public life is commendable, but it must be pursued with care. The ICDR Model, as currently proposed, risks undermining public trust through its procedural flaws and lack of fairness. Our members believe that a more considered, evidence-based approach is required to achieve the intended objectives without compromising justice or practicality.

Document 8 - Professor Stephan Lewandowsky

Humanities & Social Sciences Communications

ARTICLE

Check for updates

https://doi.org/10.1057/s41599-024-03503-6 OPEN

Liars know they are lying: differentiating disinformation from disagreement

Stephan Lewandowsky[1,2]✉, Ullrich K. H. Ecker[3], John Cook[4], Sander van der Linden[5], Jon Roozenbeek[6], Naomi Oreskes[7] & Lee C. McIntyre[8]

Mis- and disinformation pose substantial societal challenges, and have thus become the focus of a substantive field of research. However, the field of misinformation research has recently come under scrutiny on two fronts. First, a political response has emerged, claiming that misinformation research aims to censor conservative voices. Second, some scholars have questioned the utility of misinformation research altogether, arguing that misinformation is not sufficiently identifiable or widespread to warrant much concern or action. Here, we rebut these claims. We contend that the spread of misinformation—and in particular willful disinformation—is demonstrably harmful to public health, evidence-informed policymaking, and democratic processes. We also show that disinformation and outright lies can often be identified and differ from good-faith political contestation. We conclude by showing how misinformation and disinformation can be at least partially mitigated using a variety of empirically validated, rights-preserving methods that do not involve censorship.

[1] University of Bristol, Bristol, UK. [2] University of Potsdam, Potsdam, Germany. [3] University of Western Australia, Crawley, WA, Australia. [4] University of Melbourne, Melbourne, VIC, Australia. [5] University of Cambridge, Cambridge, UK. [6] Kings College London, London, UK. [7] Harvard University, Cambridge, UK. [8] Boston University, Boston, MA, USA. ✉email: stephan.lewandowsky@bristol.ac.uk

"If everybody always lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer."

— Hannah Arendt

One of the normative goods on which democracy relies is accountable representation through fair elections (Tenove, 2020). This good is at risk when public perception of the integrity of elections is significantly distorted by false or misleading information (H. Farrell and Schneier, 2018). The two most recent presidential elections in the U.S. were accompanied by a plethora of false or misleading information, which grew from false information about voting procedures in 2016 (Stapleton, 2016) to the "big lie" that the 2020 election was stolen from Donald Trump, which he and his allies have baselessly and ceaselessly repeated (Henricksen and Betz, 2023; Jacobson, 2023). Misleading or false information has always been part and parcel of political debate (Lewandowsky et al., 2017), and the public arguably accepts a certain amount of dishonesty from politicians (e.g., McGraw, 1998; Swire-Thompson et al., 2020). However, Trump's big lie differs from conventional, often accidentally disseminated, *mis*information by being a deliberate attempt to *dis*inform the public.

Scholars tend to think of disinformation as a type of misinformation and technically that is true: intentional falsehoods are but one subset of falsehoods (Lewandowsky et al., 2013) and intentionality does not affect how people's cognitive apparatus processes the information (e.g., L. K. Fazio et al., 2015). But given the real-world risks that disinformation poses for democracy (Lewandowsky et al., 2023), we think it is important to be clear at the outset whether we are dealing with a mistake versus a lie.

The tobacco industry's 50-year-long campaign of disinformation about the health risks from smoking is a classic case of deliberate deception and has been recognized as such by the U.S. Federal Courts (Smith et al., 2011, see also Civil Action 99-2496(GK) United States District Court, District of Columbia. United States v. Philip Morris Inc.). This article focuses primarily on the nature of disinformation and how it can be identified, and places it into the contemporary societal context. Wherever we make a broader point about the prevalence of false information, its identifiability or its effects, we use the term misinformation to indicate that intentionality is secondary or unknown.

An analysis of mis- and disinformation cannot be complete without also considering the role of the audience, in particular when people share information with others, where the distinction between mis- and disinformation becomes more fluid. In most instances, when people share information, they do so based on the justifiable default expectation that it is true (Grice, 1975). However, occasionally people also share information that they know to be false, a phenomenon known as "participatory propaganda" (e.g., Lewandowsky, 2022; Wanless and Berk, 2019). One factor that may underlie participatory propaganda is the social utility that persons can derive from beliefs, even if they are false, which may stimulate them into rationalizing belief in falsehoods (Williams, 2022). The converse may also occur, where members of the public accurately report an experience, which is then taken up by others, usually political operatives or elites, and redeployed for a malign purpose. For example, technical problems with some voting machines in Arizona in 2022 were seized on by Trump and his allies as being an attempt to disenfranchise conservative voters (Reid, 2022). Both cases underscore the importance of audience involvement and the reverberating feedback loops between political actors and the public which can often amplify and extend the reach of intentional disinformation (Starbird et al., 2023; Vosoughi et al., 2018), and which can often involve non-epistemic but nonetheless rational choices (Williams, 2021, 2022).

The circular and mutually reinforcing relationship between political actors and the public was a particularly pernicious aspect of the rhetoric associated with Trump's big lie (for a detailed analysis, see Starbird et al., 2023). During the joint session of Congress to certify the election on 6 January 2021, politicians speaking in support of Donald Trump and his unsubstantiated claims about election irregularities appealed not to evidence or facts but to public opinion. For example, Senator Ted Cruz cited a poll result that 39% of the public believed the election had been "rigged". Similarly, Representative Jim Jordan (R-Ohio), who is now Chairman of the House Judiciary Committee, argued against certification of the election by arguing that "80 million of our fellow citizens, Republicans and Democrats, have doubts about this election; and 60 million people, 60 million Americans think it was stolen" (Salek, 2023). The appeal to public opinion to buttress false claims is cynical in light of the fact that public opinion was the result of systematic disinformation in the first place. While nearly 75% of Republicans considered the election result legitimate on election day, this share dropped to around 40% within a few days (Arceneaux and Truex, 2022), coinciding with the period during which Trump ramped up his false claims about the election being stolen. By December 2020, 28% of American conservatives did not support a peaceful transfer of power (Weinschenk et al., 2021), perhaps the most important bedrock of democracy. Among liberals, by contrast, this attitude was far more marginal (3%).

Public opinion has shifted remarkably little since the election. In August 2023, nearly 70% of Republican voters continued to question the legitimacy of President Biden's electoral win in 2020. More than half of those who questioned Biden's win believed that there was solid evidence proving that the election was not legitimate (Agiesta and Edwards-Levy, 2023). However, the purported evidence marshaled in support of this view has been repeatedly shown to be false (Canon and Sherman, 2021; Eggers et al., 2021; Grofman and Cervas, 2023). [1] It is particularly striking that high levels of false election beliefs are found even under conditions known to reduce "expressive responding"—that is, responses that express support for a position but do not reflect true belief (Graham and Yair, 2023).

The entrenchment of the big lie erodes the core of American democracy and puts pressure on Republican politicians to cater to antidemocratic forces (Arceneaux and Truex, 2022; Jacobson, 2021, 2023). It has demonstrably decreased trust in the electoral system (Berlinski et al., 2021), and a violent constitutional crisis has been identified as a "tail risk" for the United States in 2024 (McLauchlin, 2023). Similar crises in which right-wing authoritarian movements are dismantling democratic institutions and safeguards have found traction in many countries around the world including liberal democracies (Cooley and Nexon, 2022).

In this context, it is worth noting that the situation in other countries, notably in the Global South, may differ from the situation in the U.S. (Badrinathan and Chauchard, 2024). On the one hand, low state capacity and infrastructure constraints may curtail the ability of powerful actors to spread disinformation and propaganda (though see Kellow and Steeves, 1998; Li, 2004, for discussion of the role of government-adjacent radio station RTLM in facilitating the 1994 Rwandan genocide). On the other hand, such spread can be facilitated by the fact that closed, encrypted social-media channels are particularly popular in the Global South, sometimes providing an alternative source of news when broadcast channels and other conventional media have limited reach. In those cases, dissemination strategies will also be less direct, relying more on distributed "cyber-armies" than direct one-to-millions broadcasts such as Trump's

social-media posts (Badrinathan, 2021; Jalli and Idris, 2019). The harm that can be caused by such distributed systems was vividly illustrated by the false rumors about child kidnapers shared in Indian WhatsApp groups in 2018, which incited at least 16 mob lynchings, causing the deaths of 29 innocent people (Dixit and Mac, 2018). The ensuing interplay between the attempts of the Indian government to hold WhatsApp accountable and Meta, the platform's owner, highlights the limited power that governments in the Global South hold over multinational technology corporations (Arun, 2019). As a result, many platforms do not even have moderation tools for problematic content in popular non-Western languages (Shahid and Vashistha, 2023).

The power asymmetry between corporations and the Global South has been noted repeatedly, and recent calls for action include the idea of collective action by countries in the Global South to insist on regulation of platforms (Takhshid, 2021). We have only scratched the surface of a big global issue that is in urgent need of being addressed.

Despite these differences between the Global North and South, beliefs in political misinformation can be pervasive regardless of regime type or development level (e.g., for a discussion in the context of the "developing democracy" of Brazil, see Dourado and Salgado, 2021; Pereira et al., 2022).

The political landscape of disinformation

Given that the 2020 election was lost by the Republican candidate, the finding that conservatives are more likely than liberals to believe false election claims is explainable on the basis of motivated cognition and the general finding that conspiracy theories "are for losers" (Uscinski and Parent, 2014); that is, they provide an explanation—even if only a chimerical one—for a political setback to the losing parties. There is no a priori reason to assume that susceptibility to disinformation is skewed across the political spectrum.

However, a large body of recent research on the American public and U.S. political actors has consistently identified a pervasive ideological asymmetry, with conservatives and people from the populist right being far more likely to consume, share, and believe false information than their liberal counterparts (Benkler et al., 2018; Garrett and Bond, 2021; González-Bailón et al., 2023; Grinberg et al., 2019; Guess et al., 2020a; Guess et al., 2020b; Guess et al., 2019; Ognyanova et al., 2020). Research into the asymmetry culminated in a recent analysis of the news diet of 208 million Facebook users in the U.S., which discovered that a substantial segment of the news ecosystem is consumed exclusively by conservatives and that most misinformation exists within this ideological bubble (González-Bailón et al., 2023). Although the reasons for this asymmetry are not fully understood, Lasser et al. (2022) recently showed that it also held for politicians, with Republican members of Congress disseminating far more low-quality information on Twitter/X than their Democratic counterparts. Greene (2024) reported a parallel analysis for Facebook and found the same asymmetry between politicians of the two major parties. Similarly, Benkler et al. (2018) showed how the particular structure of the American media scene, with a dense interconnected cluster of right-wing sources that is separate from the remaining mainstream, fosters political asymmetry in the use and consumption of disinformation.

This asymmetry extends beyond the political domain to health-related information, which might at first glance appear to be of sufficient importance for most people to cast aside their political leanings. A recent systematic review discovered eight studies that identified conservatism as a predictor of susceptibility to health misinformation, seven studies that found no association involving political leanings, and not a single study that showed liberals to be more misinformed on health topics than conservatives (Nan et al., 2022). The observed political asymmetry is also not limited to survey results or other behavioral measures. Wallace et al. (2023) examined vaccination and mortality data from two U.S. states (Ohio and Florida) during the COVID-19 pandemic and found a widening partisan gap in excess mortality. Specifically, whereas mortality rates were equal for registered Republican and Democratic voters pre-pandemic, a wide partisan gap—with excess death rates among Republicans being up to 43% greater than among Democratic voters—was observed after vaccines had become available for everyone. The gap was greatest in counties with the lowest share of vaccinated people and it almost disappeared for the most vaccinated counties. Similar results have been reported across U.S. states (Leonhardt, 2021). One explanation for these patterns invokes the frequent false statements by Republican politicians and conservative news networks—foremost Fox News—that discredited the COVID-19 vaccines (Hotez, 2023). In support, consumption of Fox News has been causally linked to lower vaccination rates (Pinna et al., 2022).

Moreover, a recent analysis identified a specific "Trump effect" such that even conditional on the Republican vote share, support for Trump was additionally and causally associated with a lower vaccination rate (Jung and Lee, 2023).

The political asymmetry surrounding the dissemination and consumption of misinformation must be caveated in two ways. First, although the asymmetry is substantial and pervasive it is not absolute. For some materials, such as specific conspiracy theories, the asymmetry is found to be attenuated in some studies (A. Enders et al., 2022; M. Enders and Uscinski, 2021). Second, the asymmetry observed among American politicians does not necessarily hold in other countries. Lasser et al. (2022) examined tweets by British and German parliamentarians and showed that with the exception of the extreme right in Germany (the AfD party), politicians across the mainstream spectrum were equally judicious in what information they shared in their tweets. This finding suggests that it is not conservatism per se that is associated with asymmetric reliance on misinformation, but the specific manifestation of conservatism currently dominant in the American political landscape.

Notwithstanding those caveats, the political asymmetry surrounding the dissemination and consumption of misinformation in the U.S. has been accompanied by at least two major issues: First, there has been a strong political response by Republicans in Congress who have commenced a campaign against misinformation research and researchers, claiming that the research seeks to censor conservative voices. Second, the political backlash has coincided with growing self-reflection and critique among scholars, some of whom began to question the misinformation research effort, culminating in claims that misinformation may not be sufficiently identifiable or widespread to warrant concern or countermeasures. We now take up these two issues in turn.

The politicization of misinformation research

At the time of this writing, Representative Jim Jordan, R-Ohio, has been leading a campaign against misinformation research and misinformation researchers in his role as Chairman of the House Judiciary Committee. The core allegation by Jordan and his allies[2] is that misinformation researchers are part of a purported "Censorship Industrial Complex" that is assisting the Biden administration in its purported endeavor to pressure platforms into suppressing conservative viewpoints (U.S. House of Representatives Judiciary Committee, 2023). The allegation is, however, problematic for at least four reasons: it rests on false assertions; it ironically denies first-amendment rights to researchers; it rests on a basic premise that is false; and it misunderstands the role of platforms in content moderation.

Concerning the first point, Jordan has subpoenaed several prominent academics engaged in the study of mis- and disinformation based on false assertions. For example, Dr Kate Starbird, an expert on disinformation from the University of Washington, was called to testify before Jordan's subcommittee and had to defend herself against accusations that she was colluding with the Biden administration in an effort to chill conservative speech (Nix and Menn, 2023). Core to the specific allegations against Starbird and her colleagues is a claim—initially voiced by online conspiracy theorists—that they colluded with the Department of Homeland Security to censor 22 million tweets during the 2020 election campaign. In actual fact, the researchers *collected* 22 million tweets for analysis, and flagged about 3000 (0.0001 of the total) for potential violations of Twitter's terms of use (Blitzer, 2023).

Second, Jordan's purported championing of free speech is difficult to reconcile with the chilling effect the House Committee's actions have had on the first-amendment rights of researchers. According to Starbird, "The people that benefit from the spread of disinformation have effectively silenced many of the people that would try to call them out" (Rutenberg and Myers, 2024). The deterring effect on the research community is widespread (Bernstein, 2023; Nix et al., 2023). Similarly, Facebook and YouTube have reversed their restrictions on content claiming that the 2020 election was stolen. Election disinformation, unsurprisingly, has seen an uptick in response (Rutenberg and Myers, 2024).

Third, Jordan's campaign rests on a false premise, namely that social-media platforms are biased against conservatives. Together with other conservative figures such as Tucker Carlson (formerly with Fox News) and Ben Shapiro, Jordan claimed in 2020 that "Big Tech is out to get conservatives". This claim has been shown to be wrong by several studies. For example, an analysis of Facebook engagements during the 2016 election campaign revealed that conservative outlets (Fox News, Breitbart, and Daily Caller) amassed 839 million interactions, dwarfing more centrist outlets (CNN with 191 million and ABC news with 138 million), and totaling more than the remaining seven mainstream pages in the top 10 (Barrett and Sims, 2021). Another analysis involving millions of Twitter users and 6.2 million news articles shared on the platform also found that conservatives enjoy greater algorithmic amplification than people on the political left (Huszár et al., 2022). Moreover, the Congressional January 6th Committee detailed the way in which major platforms, including Twitter and Facebook, facilitated the organization of the violent insurrection in a 122-page memo, although much of that information did not make it into the final committee report (Zakrzewski et al., 2023). Congressional investigators discovered that the platforms failed to heed their own experts' warnings about violent rhetoric on their platforms, and selectively failed to enforce existing rules to avoid antagonizing conservatives for fear of reprisals (Zakrzewski et al., 2023).

Finally, and perhaps most important, Jordan's pursuit fails to differentiate between the roles of government and the platforms, and in particular ignores the crucial role that platforms already play in shaping people's information diet (Lewandowsky et al., 2023a). In a nutshell, the internet is currently neither unregulated nor is all information on the internet equally free. Instead, nearly all content on social media is curated by algorithms that are designed to maximize dwell time in pursuit of the platforms' advertising profit (Lewandowsky and Pomerantsev, 2022; Wu, 2017). Algorithms therefore favor captivating information that keeps users engaged. Unfortunately, human attention is known to be biased towards negative information (Soroka et al., 2019), which creates an incentive for platforms to drench users in outrage-evoking content. Similar to junk food that supermarkets strategically place at checkout lanes, the information that is preferentially curated by platforms may satisfy our presumed momentary preferences while reducing our long-term well-being. If platforms were to address their role in those dynamics, for example by redesigning their algorithms, this would hardly constitute censorship. Solving a problem one has caused is good iterative design rather than bias or suppression of opinions. No one would accuse a supermarket of suppressing consumers' preferences if the checkout lanes put on offer celery instead of chocolate bars.

In summary, far from being a restorative effort in defense of free speech, Jordan's attacks are reminiscent of similar campaigns launched against inconvenient scientists by the tobacco and fossil-fuel industries (Lewandowsky et al., 2023b). In all cases scientists have been subject to personal abuse, their email correspondence is hacked or subpoenaed, and allegations are woven together from snippets of decontextualized actions or events (Blitzer, 2023). Because these attacks are systemic, the response also requires a systemic approach (Desikan et al., 2023). However, any such response seems unlikely to be achievable in the current political landscape. Scientists who work under such challenging conditions must therefore rely on other avenues to protect their integrity. The U.S. National Academy of Sciences has published a list of resources for scientists under attack.[3] Specific recommendations include responding publicly to valid criticism (without, however, engaging in a long drawn-out direct conversation with an attacker), reporting abusive messages to the authorities, and seeking support from colleagues who have been in similar situations (Kinser, 2020).

The attacks have also coincided with moves by the platforms and the courts that align with Jordan's claims. For example, the major platforms (Meta, Google, Twitter/X, and Amazon) have cut back on the number of staff dedicated to combating hate speech and misinformation (Field and Vanian, 2023). Meta (the parent company of Facebook) has been laying off employees in its "content review" team, which had been involved in countering misinformation and disinformation in the 2022 midterm election, citing confidence in improved electronic tools for detecting inauthentic accounts. It remains to be seen how the platform actions will play out during the 2024 presidential election.

In the legal arena, a Trump-appointed federal judge in Louisiana barred the Biden administration from having any contact with social-media companies and certain research institutions to discuss safeguarding elections in July 2023. The judgement echoed the claims by Jim Jordan and other Republicans that there was collusion between the White House and the social-media companies to censor conservative voices under the guise of fighting disinformation about COVID-19 during the pandemic and false election claims during the 2022 midterms. Although there are important and potentially problematic implications for free speech that arise whenever a government gets involved in managing what it considers misinformation (Neo, 2022; Vese, 2022), the Louisiana ruling was particularly broad in its prohibitions (West, 2023). The implications of the ruling include denying election officials access to information gathered by independent research bodies (the ruling lists "the Election Integrity Partnership, the Virality Project, the Stanford Internet Observatory, or any like project or group") that would enable them to debunk false election-related information and provide more accurate information instead. The Supreme Court blocked the Louisiana ruling in October 2023 (Hurley, 2023) but agreed to a full hearing later in its current term. We return to the conflict between free speech and the adverse effects of disinformation later.

The post-modern critiques of misinformation research

At the heart of research on misinformation is the belief that the concepts of truth and falsehood are essential to democracy, to cognition, and to daily life, and that the status of many, but of course not all claims, can be determined with sufficient accuracy

to warrant rebuttal of false information. For example, the "big lie" about a stolen election is just that—it is a lie with no sustainable evidentiary support and it is routinely referred to as such in the scholarly literature (e.g., Arceneaux and Truex, 2022; Canon and Sherman, 2021; Graham and Yair, 2023; Henricksen and Betz, 2023; Jacobson, 2021, 2023; Painter and Fernandes, 2022). The lie has been rejected by 62 American courts, all of which dismissed or ruled against law suits questioning the legitimacy of the election by Donald Trump or his supporters.[4]

It is curious that the reaction by Trump and some of his most ardent public supporters to such determinative judgments about the falsity of his claims has not been to claim that they are in fact true, but to attack the idea that objective knowledge is even possible. When confronted with a lie, Trump's adviser Kellyanne Conway once famously quipped that she was presenting "alternative facts." On another occasion, Trump's attorney Rudy Giuliani declared that "truth isn't truth." Such a strategy seems oddly reminiscent of the postmodernist critique of the possibility of objective knowledge, which first arose as a core aspect of 1930s fascism and was then adapted by left-wing literary criticism from the 1960s onward (Lewandowsky, 2020). At that time, humanities scholars had grown increasingly uncomfortable with the idea that facts were just facts, and that there was no role for considering the personal or political interests of those who were engaged in the pursuit of empirical knowledge. In this, postmodernists raised an important point of self-reflection for scientists and others who blithely claimed that there was an impenetrable wall between facts and values. But then they took things too far. Derrida claimed that there was no such thing as objective knowledge. Foucault went on to suggest that —given this— all knowledge claims were nothing more than an assertion of the political interests of the investigator (McIntyre, 2018, p. 124).

This led to the "science wars" of the 1990s, when scientists and their allies fought back against subjectivism and relativism to defend the importance of objective knowledge at least as a regulative ideal of empirical inquiry. This particular attack on science eventually dissipated —and in the face of damage it had done to objective knowledge claims like the reality of global warming, some postmodernists such as Bruno Latour eventually even apologized (Latour, 2004)— but the damage was already done. Meanwhile, both the corporate sector and the religious and political right wing had once again taken up the strategy in their attacks on science. The advantage of post-modernism for anti-democratic purposes is obvious, and has echoes of authoritarian attacks on truth-tellers and their defenders throughout history. Indeed, to someone who embraces the idea that their political ideology *should* have supremacy over objective reality, the advantages of postmodernism are clear. Not only can falsehoods about the economy, crime, and political violence be offered as "alternative narratives" to carefully-measured statistics or other forms of evidence, but the credibility of *any* party as an objective truth-teller can be undermined. And this suits the authoritarian just fine—for where there is no truth then there can be no blame or accountability either.

Hannah Arendt long ago recognized the dangers of this strategy when she wrote: "the ideal subject of totalitarian rule is not the convinced Nazi or the convinced communist, but people for whom the distinction between fact and fiction … true and false … no longer exist." This easy political slide into postmodernism does violence to the idea that truth matters, that facts can be discovered through empirical analysis, and that it is crucial to attempt to discern the facts before we can make good policy—especially when we hold competing values that will impact policy choice. And this is true even more so in an era when the creation and amplification of knowledge claims are so easily subject to digital manipulation and weaponization by anyone who has a personal or political interest. Fortunately, researchers have developed conceptual, cognitive, and computational tools that permit the differentiation between legitimate contestation of facts on the one hand, and misinformation and willful disinformation on the other.

The identifiability of contested facts

Notwithstanding our rejection of the postmodernist project, we do not dispute its core idea that many contested assertions cannot be unambiguously adjudicated by referring to "facts". There are indeed cases in which different actors may legitimately question each other's "facts". In our view, these ambiguous cases are precisely those that merit democratic debate and contestation. When conducted in good-faith, such debates can be particularly revealing because both sides can marshal evidence in support.

To illustrate, consider the recent controversy surrounding a machine-learning tool known as COMPAS (Dressel and Farid, 2018), which is intended to assist judges in the U.S. by predicting the likelihood of recidivism of a specific offender. Critics accused COMPAS of being racially biased based on statistical analysis of the evidence (Angwin et al., 2016). The case rested on the observation that among defendants who ultimately did not re-offend, the algorithm misclassified African-Americans as being at risk of re-offending more than twice as often as White offenders. This misclassification can have serious consequences for a person because judges are inclined to treat high-risk defendants more harshly.

Proponents of COMPAS rejected this charge and argued that the algorithm was not racially biased because it predicted recidivism equally for Black and White offenders for each of its 10 risk categories. That is, the classification into risk categories based on a large number of indicator variables was racially unbiased—a Black person's actual probability of re-offending was the same as that of a White person with the same risk score (Dieterich et al., 2016).

It turns out that it is mathematically impossible to simultaneously satisfy both forms of fairness—calibration and classification—when the base rates of re-offending differ between groups (Berk et al., 2021; Lagioia et al., 2023). That is, if a greater share of Black people are classified as high-risk—which the algorithm does in an unbiased manner—then it necessarily follows that a greater share of Black defendants *who do not re-offend* will also be mistakenly classified as high-risk. In those circumstances, it would be inappropriate to accuse one or the other side of spreading misinformation, as each party has mathematical justification for their position and a resolution can only be attained through a value-laden policy discussion. Indeed, to our knowledge, the main contestants in this debate—Northpointe, the manufacturer of COMPAS (Dieterich et al., 2016) and ProPublica, a public-interest media organization (Angwin et al., 2016)—did not level charges of misinformation against each other despite engaging in robust debate.

A similar controversy with even greater stakes arose in the context of the COVID-19 vaccine rollout in the U.S. in 2021. Unlike most other countries, which vaccinated their populations according to age alone—with the elderly being given highest priority because of their much higher mortality rate from COVID-19—the U.S. Advisory Committee on Immunization Practices (ACIP) favored a policy that gave higher priority to essential workers (e.g., food and transport workers) than the elderly. This policy was partially motivated by the fact that racial minorities (Blacks and Hispanics) are underrepresented among adults over 65, whereas they are slightly over-represented among essential workers—thus, under an age-based policy the share of Whites who receive the vaccine would have initially been greater

than their proportion in the population would have warranted. Conversely, Blacks would have been underrepresented among the vaccinated early on (Mounk, 2023). This inequity could be avoided by first vaccinating essential workers among whom racial minorities were over-represented. However, because the age distribution of essential workers has a much lower average, fewer lives were saved among vaccinated essential workers—whose young age rendered their risk of dying from COVID-19 low to begin with—than would have been saved among the elderly had they been vaccinated (Rumpler et al., 2023). Modeling has confirmed that while the essential-worker policy introduced racial equity in terms of doses administered, more lives would have been saved *in all ethnic groups* under an age-based policy (Rumpler et al., 2023). Again, the apparent fairness of a policy depended on the outcome measure: doses administered vs. lives saved. Given the unequal distributions of different ethnic and racial groups across different ages, no mathematical possibility exists to settle on a single "fair" policy. Public opinion appears to have been broadly in line with the policy ultimately adopted by ACIP (Persad et al., 2021).

The controversies surrounding COMPAS and ACIP's vaccination policy are just two instances of a much wider problem, which is that when issues become sufficiently complex, even good-faith actors may find it impossible to agree. One reason is that cognitive limitations prevent a full Bayesian representation (the gold standard of rationality) of the problem (Pothos et al., 2021). Instead, people are forced to simplify their representations, for example by partitioning their knowledge (Lewandowsky et al., 2002). Persistent and irresolvable disagreements are thus almost ensured by human cognitive limitations (Pothos et al., 2021). The second reason is that people differ in their values and weigh evidence differently even if all parties can agree on underlying facts (Walasek and Brown, 2023).

Nonetheless, controversies such as those surrounding COMPAS and ACIP's vaccination policy do not give licence to political actors to obscure the debate through falsehoods, misleading claims, or lies. On the contrary, proper debate of those issues is only possible in the absence of falsehoods because their resolution ultimately requires a trade-off of values that is best arrived at by weighing the importance of different competing sets of evidence. We therefore reject recent academic voices that have questioned whether misinformation can be reliably identified at all (Acerbi et al., 2022; Adams et al., 2023; Harris, 2022; van Doorn, 2023; Yee, 2023a, 2023b). We suggest that its identification is essential and, as we show next, empirically well supported.

The identifiability of misinformation

We place our case into the context of the more extreme end of the academic critique because it involves positions that are antithetical to ours, calling into question the entire idea of fact-checking. For example, Uscinski (2015) raised the specter that fact-checking is merely a "veiled continuation of politics by means of journalism" (p. 243). Yee (2023a) argued more broadly that any deference to "epistemic elites"—including not only fact-checkers but also academics, researchers, or journalists—is problematic, and assessment of the quality of information should include democratic elements "that are participatory, transparent, and fully negotiable by average citizens" (Yee, 2023a, p. 1111). This demand has several problematic implications. First, it does not explain who counts as "average citizen" and who would belong to the "elite". At what point should individuals seeking to counter misinformation begin to recuse themselves for fear of accidentally treading on "average" citizens? Is a virologist too "elite" to correct misinformation surrounding the origin of a new virus? What about citizens with a PhD or Master's degree, how are they being classified? Second, why exactly would one exclude epistemic elites, such as investigative journalists or forensic IT experts, from identifying bad-faith actors such as foreign "bots" or "trolls"? Are average citizens really better at this task than network scientists? Should we decide by social-media poll whether a new strain of avian flu is contagious to humans (Lewandowsky et al., 2017)? Probably not. There are obviously many domains that benefit from expert assessment of claims.

Nonetheless, there has been much research that has revealed the competence of crowds in the context of fact-checking. For example, Pennycook and Rand (2019) showed that crowdsourced trust ratings of media outlets were quite successful in the aggregate when compared to ratings by professionals, notwithstanding substantial partisan differences. This basic finding has been replicated and extended several times (M. R. Allen et al., 2024; Martel et al., 2024), with community-based fact-checking of COVID-19 content being 97% accurate in one study (M. R. Allen et al., 2024). Care must, however, be taken that crowds are politically balanced. When people can choose what content to evaluate, as in Twitter/X's crowdsourced "Birdwatch" fact-checking program (now known as Community Notes), partisan differences among contributors may limit the value of the crowdsourcing (J. Allen et al., 2022). The crowdsourcing results show not only that average citizens can match the competence of experts in the aggregate, but they also reaffirm that misinformation is identifiable.

Much recent research has uncovered specific "fingerprints" that can enable people as well as machines to infer the likely quality or accuracy of content. Misinformation has been shown to be suffused with emotions, logical fallacies, and conspiratorial reasoning (Blassnig et al., 2019; Carrasco-Farré, 2022; Fong et al., 2021; Musi et al., 2022; Musi and Reed, 2022). For example, critical thinking methods offer a qualitative approach to deconstructing arguments in order to identify the presence of reasoning fallacies (Cook et al., 2018).

Quantitatively, one study found that compared to reliable information, misinformation is less cognitively complex and 10 times more likely to rely on negative emotional appeals (Carrasco-Farré, 2022). In confirmation, numerous other studies show that misinformation is, on average, more emotional than factual information (for a systematic review, see Peng et al., 2023) Upward of 75% of anti-vaccination websites use negative emotional appeals (Bean, 2011) and linguistic analyses show that conspiracy theorists use significantly more fear-driven language as compared to scientists (Fong et al., 2021).

Emotion also plays a role in the receivers' behavior. People have been shown to be more susceptible to misinformation when put in an emotional state (Martel et al., 2020), which helps explain the preferential and more rapid diffusion of unreliable versus reliable information online (Pröllochs et al., 2021; Vosoughi et al., 2018).

Critics may argue that the datasets used for determining what constitutes "misinformation" and "reliable" information are limited or biased or that the mere prevalence of these cues is not evidence of their diagnosticity in real-world contexts. However, computational machine-learning work relying on a large variety of different URL sources and fact-checked datasets has confirmed that the results are robust and generalizable (Ghanem et al., 2020; Kumari et al., 2022; Lebernegg et al., 2024). A recent comprehensive study which combined many of the available cues found that they have high diagnostic and predictive validity and help discriminate between false and true information, with state-of-the-art models reaching over 83% classification accuracy (Lebernegg et al., 2024). Moreover, real-world training on fake news detection, such as logical fallacy training, helps people accurately discriminate between misleading and credible news

(e.g., Hruschka and Appel, 2023; Lu et al., 2023; Roozenbeek et al., 2022).

In summary, the available evidence shows quite convincingly that misinformation can be identified by both humans and machines with considerable accuracy. As we show next, we can go beyond mere identification as there are also at least three ways in which one can ascertain the deceptive intent underlying disinformation if present. Identification of deceptive intent is particularly pertinent because it allows information to be safely discounted without requiring a detailed analysis of its factual status.

The identifiability of willful disinformation

For decades, the hallmark of Western news coverage about politicians' false or misleading claims was an array of circumlocutions that carefully avoided the charge of lying—that is, knowingly telling an untruth with intent to deceive (Lackey, 2013)—and instead used adverbs such as "falsely", "wrongly", "bogus", or "baseless" when describing a politician's speech. Other choice phrases referred to "unverified claims" or "repeatedly debunked claims". This changed in late 2016, when the *New York Times* first used the word "lie" to characterize an utterance by Donald Trump (Borchers, 2016). The paper again referred to Donald Trump's lies within days of the inauguration in January 2017 (Barry, 2017) and it has grown into a routine part of its coverage from then on. Many other mainstream news organizations soon followed suit and it has now become widely accepted practice to refer to Trump's lies as lies.

Given that lying involves the intentional uttering of false statements, what tools are at our disposal to infer a person's intention when they utter falsehoods? How can we know a person is lying rather than being confused? How can we infer intentionality?

Anecdotally, defenders of Donald Trump's lies have raised precisely that objection to the use of the word "lie" in connection with his falsehoods. This objection runs afoul of centuries of legal scholarship and Western jurisprudence. Brown (2022) argues that inferring intentionality from the evidence is "ordinary and ubiquitous and pervades every area of the law" (p. 2). Inferring intentionality is the difference between manslaughter and murder and is at the heart of the concept of perjury—namely, willfully or knowingly making a false material declaration (Douglis, 2018).

There are at least three approaches that can be pursued to infer intentional deception by a communicating agent with varying degrees of confidence. The first approach is statistical and relies on linguistic analysis of material. Unlike people, who are not very good lie detectors despite performing (slightly) above chance (Bond and DePaulo, 2006; Mattes et al., 2023), recent advances in natural language processing (NLP) have given rise to machine-learning models that can classify texts as deceptive or honest based on subtle linguistic clues (e.g., Braun et al., 2015; Davis and Sinnreich, 2020; Van Der Zee et al., 2021). To illustrate, a model that relied on analysis of the distribution of different types of words achieved 67% accuracy (considerably better than the 52% achieved by human judges) on texts generated by speakers who were either instructed to lie or to be honest. Using the same analysis approach, Davis and Sinnreich (2020) trained a model to classify tweets by Donald Trump as true or false by using independent fact-checks as ground truth. The model was able to classify tweets with more than 90% accuracy, suggesting that Trump uses subtly different language (e.g., more negative emotion, more prepositions and discrepancies) when communicating untruths. A similar model of Trump's tweets was developed by Van Der Zee et al. (2021), who additionally applied 26 extant models from the literature to Trump's tweets and showed that most of them performed above chance despite being developed on very different materials. In summary, NLP-based approaches have repeatedly shown their value in the classification of speech into honest and deceptive. The fact that those models succeed also when applied to the tweets of Donald Trump implies at the very least that Trump's falsehoods are not uttered at random or accidentally but are deployed using specific linguistic techniques.

In general, machine-learning approaches to deception detection have shown promise. A recent systematic review identified 81 studies, 19 of which achieved accuracies in excess of 90%, with a further 15 exceeding 80% accuracy (Constâncio et al., 2023). The machine-learning models in that ensemble were trained on a variety of corpora, ranging from reviews on Tripadvisor (either true or generated with the intent to deceive; Barsever et al., 2020) to segments of a radio game show dedicated to bluff detection by the audience (Papantoniou et al., 2021). In all cases, the ground truth (i.e., whether or not deceptive intent was present) was unambiguously known, and the models learned to identify deceptive text based on linguistic analysis with considerable albeit imperfect success.

The second approach to establish willful deception relies on analysis of internal documents of institutions such as governments or corporations. Comparison of the internal knowledge to public stances of the same entities can identify active deception, especially when it is large-scale. Numerous such cases exist, mainly involving corporations and their associated infrastructure such as think-tanks and other front groups (Ceccarelli, 2011; Oreskes and Conway, 2010). For example, as early as the 1920s, the electricity industry organized a propaganda campaign to falsely insist that private sector electricity was cheaper and more reliable than electricity generated in the public sector (Oreskes and Conway, 2023). The tobacco industry's activities to mislead the public about the dangers from smoking are well documented and established beyond reasonable doubt (e.g., Cataldo et al., 2010; Fallin et al., 2013; Francey and Chapman, 2000; Proctor, 2012). The tobacco industry was well aware of the link between smoking and lung cancer in the 1950s and 1960s (Proctor, 2012), and yet continued publicly to dispute that medical fact using a variety of propagandistic means (Landman and Glantz, 2009; Proctor, 2011). Similarly, analysis of internal documents of the fossil-fuel industry has revealed that industry leaders, in particular ExxonMobil, were fully aware of the reality of climate change and its underlying causes (Supran and Oreskes, 2017, 2021) while simultaneously expending large sums to deny its existence in public (J. Farrell, 2016) and to prevent Congress from enacting climate-mitigation legislation (Brulle, 2018). Ironically, ExxonMobil's scientists projected global temperatures in the 1970s and 1980s with comparable skill as independent academics at the time (Supran et al., 2023). As Baker and Oreskes (2017) argued, the best explanation for ExxonMobil's conduct is that they knowingly deceived the public by funding a disinformation machine that denied the realities of climate change. This approach admittedly requires considerable resources and skill, and it is comparatively slow, but in exchange the results it yields are particularly diagnostic and demonstrably useful in litigation. In the case of the tobacco industry, this was the basis for a conviction of Phillip Morris under federal racketeering (RICO) law. The appeals in that case explicitly noted that Phillip Morris intentionally deceived the public and that first-amendment (free speech) rights did not apply as they do not protect fraud or deliberate misrepresentation (Farber et al., 2018). In the case of the fossil fuel industry, litigation has not been met with notable success at the time of this writing, but the "Exxon knew" campaign, based on research by Supran and colleagues (Supran et al., 2023; Supran and Oreskes, 2017, 2021), has had considerable public impact with 178 relevant media articles identified by Google News.[5]

The final approach to identifying intentional deception resembles the approach involving institutional documents but specifically focuses on lies promulgated by identifiable individuals. We illustrate this approach with Donald Trump's big lie about the 2020 presidential elections, focusing on statements made in courts of law. Although Trump was making widespread public accusations of fraud, his lawyers—who filed more than 60 lawsuits in connection with the election—did not echo those accusations in court. Quite on the contrary, his lawyers frequently disavowed any mention of fraud in court despite their very different public stance. For example, Rudy Giuliani, one of Trump's lead attorneys, stood outside a landscaping business on the day most networks declared the election for Biden, and thundered that "It's [the election] a fraud, an absolute fraud." Ten days later, being questioned by a federal judge in Pennsylvania during one of Trump's lawsuits (dealing with whether local election officials in Pennsylvania should have allowed voters to fix problems with their mail-in ballots after submitting them), he declared "This is not a fraud case" (Lerer, 2020). This pattern was pervasive: Trump's lawyers continued to back away from suggestions that the election was stolen and admitted in court that there was no evidence of fraud, all in contradiction to their plaintiff's public statements (Lerer, 2020).

Notwithstanding the careful hedging of their claims in court, the frivolous suits filed on behalf of Trump resulted in sanctions for several of his attorneys. Two lawyers who did claim widespread voter fraud not only had their suit dismissed but were also sanctioned $187,000 by a federal judge in Colorado for their frivolous, meritless case (Polantz, 2021). The decision was upheld on appeal, and the Supreme Court declined to hear a further appeal by the lawyers (Scarcella, 2023). Altogether, 22 Trump lawyers have been identified who face sanctions in litigation, criminal prosecutions, and state bar disciplinary proceedings. In all cases, what appears to be at issue is violation of the Model Code of Conduct, in particular rules stipulating that claims must be meritorious and that lawyers must exhibit candor and truthfulness (Neff and Fredrickson, 2023).

Since the flurry of lawsuits in late 2020, Trump lawyer Sidney Powell has pleaded guilty to charges arising from her involvement in pushing the big lie. Ms Powell pleaded guilty to "conspiracy to commit intentional interference with performance of election duties" and agreed to cooperate with prosecutors in a criminal case against Donald Trump (Fausset and Hakim, 2023). Two further Trump lawyers have pleaded guilty in the same case and agreed to testify truthfully about other defendants (Blake, 2023).

In a civil suit brought against Rudy Giuliani by two election workers in Georgia, whom he had publicly accused of election fraud, Giuliani conceded before trial that those statements were false (Brumback, 2023). The election workers were awarded $148 million in damages, causing Giuliani to file for bankruptcy in late 2023 (Aratani and Oladipo, 2023). In a further twist, Giuliani repeated his false claims during the trial *outside the court room* even while his lawyers conceded in court that they were wrong (Hsu and Weiner, 2023).

Giuliani was promptly sued again by the election workers, and at the time of this writing the suit was still under way (Hsu and Weiner, 2023).

The big lie was not just curated and pushed by politicians seeking to cling to power and their attorneys. It is now public knowledge that one major news network, Rupert Murdoch's Fox News, knowingly amplified claims about the election that network executives knew to be false. The fact that Fox lied became apparent during a defamation suit filed by Dominion Voting Machines against the network over false allegations that the voting machines had been rigged to steal the 2020 election. As trial was about to begin, Fox News agreed to pay Dominion $787.5 million and acknowledged that the network had broadcast false statements. The discovery process that preceded trial had uncovered numerous documents and emails that revealed that senior network executives and hosts were convinced that the allegations about the election made by Trump and his allies were untrue (e.g., Peltz, 2023; Terkel et al., 2023). The network continued to air those allegations and its CEO instructed staff that fact-checking "had to stop" because it was bad for business (Levine, 2023). One scholar put it succinctly: "Fox News deliberately misleads the audience for profit" (Nyberg, 2023, p. 1). Although Fox has been repeatedly implicated in spreading disinformation with harmful consequences for the American public (Ash et al., 2023; Bolin and Hamilton, 2018; Bursztyn et al., 2020; DellaVigna and Kaplan, 2007; Feldman et al., 2012; Kull et al., 2003; Simonov et al., 2022), the Dominion case provided a unique opportunity to ascertain that, at least in this case, the network was knowingly lying to its audience.

The preceding examples illustrate the approaches available to establish—with some degree of confidence—the intention to deceive that is the core element of lies. Our examples are not intended to be exhaustive but they illustrate the options available to researchers, journalists, and the public to uncover when they are being lied to. The examples also put to rest several generous auxiliary assumptions that have been made about lies in politics, such as their presumed inevitability because issues can be so nuanced that complete honesty is impossible. Contrary to that assumption, the fact that a person's rhetoric can differ strikingly between courts of law—where penalties apply for misrepresentations and perjury—and politics—where accountability is notoriously absent—not only reveals the intention to deceive but also the person's sensitivity to the consequences of their speech.

We have already noted that the contrast between what companies such as ExxonMobil or Philip Morris said in public about their products and what they discussed in private was sufficient to provoke legal consequences. Similar arguments, that fraudulent political speech should not be protected by the First Amendment, have been advanced in the context of Trump's big lie (Henricksen and Betz, 2023).

Although our examination was necessarily limited to a small number of cases, they suffice to illustrate a pathway towards pinpointing intentional disinformation by analysing the utterances of the liars themselves, be they corporations, politicians, or media organizations. We believe that the basic approach is of considerable generality, extending to numerous recorded instances:

- Politicians catching themselves lying by changing their story, indicating they were telling an untruth on at least one of those occasions (O'Toole, 2022, p. 427).
- Attorneys of conspiracy theorist Alex Jones—who was sued for his claims that the Sandy Hook massacre never happened by parents of the victims—seeking to defend him by calling him a performance artist who should not be taken seriously (Borchers, 2017).
- Alex Jones himself admitting in court that the Sandy Hook shooting was "100% real" after having misled millions of people for many years (Associated Press, 2022).
- Fox News requiring their employees to be vaccinated against COVID-19 or submit to daily testing while the network routinely broadcast anti-vaccination content (Darcy, 2021).
- Tucker Carlson, former Fox News host, openly admitting that he lies on air (Muzaffar, 2021).

Moving forward

Our work explored three fundamental premises: First, that democracy rests on a foundation of common knowledge (H. Farrell and Schneier, 2018) and that it is imperiled if citizens

cannot agree on basic facts such as the integrity of elections (H. Farrell and Schneier, 2018; Tenove, 2020). Second, that while democratic debate—including evidence-informed policy-making—often involves contestation of facts (e.g., Kuklinski et al., 1998), this does not licence the use of outright lies and propaganda to willfully mislead the public (Lewandowsky, 2020). Third, that it is often possible to identify falsehoods, disinformation, and lies and differentiate them from good-faith political and policy-related argumentation.

At the time of this writing, Donald Trump is the Republican nominee for the 2024 presidential election. His campaign has rolled out an explicitly authoritarian agenda for his second term (Arnsdorf and Stein, 2023). The authoritarian agenda is likely to result in less free speech, rather than more, which is ironic in light of the fact that people such as Jim Jordan, who are attacking the idea of studying disinformation, do so under the banner of defending the First Amendment. Against this background, the question of how to address Donald Trump's lies in particular and misinformation in general takes on particular importance.

At the more pessimistic end, Barkho (2023) posed three questions about the success of fact-checking Trump's claims: first, have fact-checkers succeeded in persuading Trump to stop disseminating lies? Second, have the long inventories of falsehoods compiled by fact-checkers embarrassed or shamed Trump? Third, has fact-checking changed public perception of what constitutes truth? At first glance, the answer to all three questions might appear to be a resounding "no" (even though the counterfactual is, of course, unknown). However, at the more optimistic end of the spectrum, experimental studies in which election-fraud misinformation was corrected have found positive effects on trust in electoral processes (Bailard et al., 2022; Painter and Fernandes, 2022), including among Republican respondents and supporters of Trump. Those findings should give rise to a sliver of optimism that even partisans are receptive to corrective messages about election integrity, and therefore underscore the value of disinformation research.

Correcting lies about elections is arguably compatible with the spirit of a democracy. But what is the democratic legitimacy of broader countermeasures against misinformation and disinformation? It is straightforward to explore techniques with which to correct misconceptions in an experiment, in particular if the misinformation is introduced in the experiment itself (e.g., Ecker et al., 2011). It is less straightforward to deploy such techniques in the public sphere. Who determines what is "misinformation", and what is "correct"? And how narrow is the gap between correcting misinformation and banning it? Several countries have recently outlawed "fake news" (e.g., Burkina Faso, Cambodia, Hungary, India, Malaysia, Singapore, and Vietnam) whose democratic credentials are at best questionable. In those cases, fake news can damage democracy not only by disinforming the public but also because countermeasures can be used to curb civil liberties and justify authoritarian crackdowns (Neo, 2022; Vese, 2022). Indeed, given that Donald Trump has routinely labeled any media coverage he did not like as "fake news", perhaps the worst response to misinformation would be a law against fake news designed by Donald Trump and his allies.

There are, however, numerous ways in which the public can be better protected by the platforms—in particular if prodded into action by suitable regulations—against disinformation. One avenue involves content moderation and removal of unacceptable or problematic content, such as hate speech. The public is broadly supportive of moderation in certain cases (Kozyreva, Herzog, et al., 2023), and the European Union's recent Digital Services Act (DSA) acknowledges a role for content moderation while highlighting the need for transparency of the underlying rules (for details, see Kozyreva, Smillie, et al., 2023). In addition, there are a number of alternative approaches that aim to inform or educate consumers rather than govern content directly. Those approaches have the advantage that they side-step concerns about censorship and that they are demonstrably scalable and readily deployable by the platforms.

One avenue involves the provision of "nutrition labels", that is, indicators of the quality of a source. Reliable indicators of quality exist that are based on basic journalistic principles (Lin et al., 2023), and it is well-known that perceived source credibility can influence misinformation persuasiveness (Nadarevic et al., 2020; Prike et al., 2024). The effectiveness of source-quality indicators can be enhanced by introducing friction, for example, by requiring users to expend additional clicks to make information visible (L. Fazio, 2020; Pillai and Fazio, 2023). Naturally, such indicators cannot be perfect, and even sources of widely-acknowledged high quality can also publish dubious content. This makes it important to go beyond credibility and consider alternative approaches, such as those that boost users' ability to spot deception and enhance their information-discernment skills. This can range from teaching "critical ignoring" (Kozyreva, Wineburg, et al., 2023), which enables people to ignore information that is unlikely to warrant expenditure of our limited attention, to psychological inoculation or "prebunking" (Lewandowsky and van der Linden, 2021; Roozenbeek et al., 2022), which involves refuting a lie in advance by explaining the rhetorical techniques that disinformers use to mislead consumers (e.g., scapegoating, false dichotomies, ad hominem attacks, and so on). Through short "edutainment" videos that are displayed as ads or public-service messages, this approach has been scaled on social media to empower millions of people to spot manipulation techniques (Goldberg, 2023). Meta-analyses have affirmed the efficacy of the inoculation approach (Banas and Rains, 2010; Lu et al., 2023). However, while standard debunking and prebunking interventions promise to be effective regardless of the cultural context in which they are applied (Blair et al., 2024; Pereira et al., 2023; Porter and Wood, 2021; but see Pereira et al., 2022), the effects of other interventions such as media-literacy training may be less robust in the Global South (Badrinathan, 2021). Some interventions developed and successfully applied in the Global North may also be less suitable in less-developed countries, if for example they target dissemination channels that have limited relevance locally (Badrinathan and Chauchard, 2024; de Freitas Melo et al., 2019).

Overall, much is now known about various cognitively-inspired countermeasures to correct misinformation or to protect people against being misled in the first place. For further extensive discussion of these countermeasures, see Ecker et al. (2022) and Kozyreva et al. (2024). Some of the cognitive science of misinformation has been reflected in European regulatory initiatives, such as the strengthened Code of Practice on Disinformation (Kozyreva, Smillie, et al., 2023). In addition, specific evidence-based recommendations for platforms have been developed by Roozenbeek et al. (2023) and Wardle and Derakhshan (2017).

Our work has also identified several important questions for future research. We consider the long-term consequences of misinformation on society to be a particularly pressing issue. We have a reasonably good understanding of the individual-level cognitive processes that are engaged when a person is exposed to a single piece of misinformation (Ecker et al., 2022). We know very little about the cognitive and social consequences for an individual who is inundated with information of dubious quality for prolonged periods of time. We do not know how societies are affected by epistemic uncertainty and chaos in the long run. Numerous indicators suggest that Western societies, in particular the United States, are

ailing (e.g., Lewandowsky et al., 2017), but the attribution of those trends to misinformation or epistemic chaos is difficult. On those occasions where researchers have successfully isolated causal effects, they tend to implicate certain media organs (e.g., Fox News in particular) in compromising public health (Bursztyn et al., 2020; Simonov et al., 2020), and they have identified the role of social media in causing ethnic hate crimes and xenophobia (Bursztyn et al., 2019; Müller and Schwarz, 2021). However, it is unclear as yet how generalizable those findings are and much additional work remains to be done (for a review, see Lorenz-Spreen et al., 2022).

Future research should also address some of the limitations of fact-checking, such as the difficulties of verifying statements about the future (Nieminen and Sankari, 2021) or arguments that employ the rhetorical technique of "paltering" — that is, the use of truthful statements to convey a misleading impression (Lewandowsky et al., 2016; Rogers et al., 2017). One approach is to focus on what is pragmatically useful for people to make informed decisions, such as whether a claim is misleading (Birks, 2019), with critical thinking methods offering a means of identifying the presence of logical fallacies (Cook et al., 2018).

Increasing research attention is being paid to the concept of discernment; that is, the extent to which accurate misinformation is believed more than misinformation (Pennycook and Rand, 2021). Focusing on discernment rather than acceptance of misinformation guards against inadvertently developing interventions that reduce belief in facts and misinformation equally. A general cynicism and disbelief of everything does not solve the misinformation problem. Instead, we must boost people's ability to distinguish between facts and falsehoods.

Conclusion

We began the paper with a quote from Hannah Arendt, one of the foremost analysts of 20th century totalitarianism. It is worth here revisiting the same quotation in its extended form, which underscores the urgency of finding a solution to the epistemic crisis affecting democracy in the U.S. and beyond:

> "If everybody always lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer…. And a people that no longer can believe anything cannot make up its mind. It is deprived not only of its capacity to act but also of its capacity to think and to judge. *And with such a people you can then do what you please*." (our emphasis)
>
> — Hannah Arendt

Received: 25 January 2024; Accepted: 22 July 2024;
Published online: 31 July 2024

Notes

1 Further detailed debunkings of election disinformation are provided by the Cybersecurity and Infrastructure Security Agency at https://www.cisa.gov/topics/election-security/rumor-vs-reality.

2 We focus here on the activities of Jim Jordan because he is the acknowledged leader of a political counter movement aimed at misinformation research. This must not be taken to imply that Jordan is the only political actor involved in this effort.

3 https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/DC4CDD2AC5D4B2DB08255A7EA6244AA9D7CA6F951C22?noSaveAs=1

4 One ruling that was initially in Trump's favor was later overturned by the Pennsylvania Supreme Court. Canon and Sherman (2021) provide a list of cases.

5 Search conducted on 10 April 2024.

References

Acerbi A, Altay S, Mercier H (2022) Research note: Fighting misinformation or fighting for information? Harv Kennedy School Misinform Rev 3. https://doi.org/10.37016/mr-2020-87

Adams Z, Osman M, Bechlivanidis C, Meder B (2023) (Why) Is Misinformation a Problem? Perspect Psychol Sci 17456916221141344. https://doi.org/10.1177/17456916221141344

Agiesta J, Edwards-Levy A (2023) CNN poll: Percentage of Republicans who think Biden's 2020 win was illegitimate ticks back up near 70%. CNN. https://edition.cnn.com/2023/08/03/politics/cnn-poll-republicans-think-2020-election-illegitimate/index.html

Allen J, Martel C, Rand DG (2022) Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. CHI Conf Hum Factors Comp Syst 1–19. https://doi.org/10.1145/3491102.3502040

Allen MR, Desai N, Namazi A, Leas E, Dredze M, Smith DM, Ayers JW (2024) Characteristics of X (formerly Twitter) Community Notes addressing COVID-19 vaccine misinformation. JAMA 331:1670. https://doi.org/10.1001/jama.2024.4800

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias: There's software used across the country to predict future criminals and it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Aratani L, Oladipo G (2023) Giuliani files for bankruptcy after judge rules Georgia election workers can collect $148m. The Guardian. https://www.theguardian.com/us-news/2023/dec/21/giuliani-148-million-damages-georgia-lawsuit

Arceneaux K, Truex R (2022) Donald Trump and the Lie. Perspect Polit 1–17. https://doi.org/10.1017/S1537592722000901

Arnsdorf I, Stein J (2023) Trump touts authoritarian vision for second term: 'I am your justice'. Washington Post. https://www.washingtonpost.com/elections/2023/04/21/trump-agenda-policies-2024/

Arun C (2019) On WhatsApp, rumours, and lynchings. Econ Polit Wkly 54(6):30–35

Ash E, Galletta S, Hangartner D, Margalit Y, Pinna M (2023) The effect of Fox News on health behavior during COVID-19. Polit Anal 1–10. https://doi.org/10.1017/pan.2023.21

Associated Press (2022) Alex Jones concedes that the Sandy Hook attack was '100% real'. NPR. https://www.npr.org/2022/08/03/1115414563/alex-jones-sandy-hook-case

Badrinathan S (2021) Educative interventions to combat misinformation: evidence from a field experiment in India. Am Polit Sci Rev 1–17. https://doi.org/10.1017/S0003055421000459

Badrinathan S, Chauchard S (2024) Researching and countering misinformation in the Global South. Curr Opin Psychol 55:101733. https://doi.org/10.1016/j.copsyc.2023.101733

Bailard, CS, Porter, E, & Gross, K (2022). Fact-checking Trump's election lies can improve confidence in U.S. elections: Experimental evidence. Harv Kennedy School Misinform Rev. https://doi.org/10.37016/mr-2020-109

Baker E, Oreskes N (2017) Science as a game, marketplace or both: a reply to Steve Fuller. Soc Epistemol Rev Reply Collect 6:65–69

Banas JA, Rains SA (2010) A meta-analysis of research on inoculation theory. Commun Monogr 77:281–311

Barkho L (2023) A critical inquiry into US media's fact-checking and compendiums of Donald Trump's falsehoods and "lies". In A Akande (Ed.) The perils of populism: The end of the American century? (pp. 259–278). Springer

Barrett PM, Sims JG (2021) False accusation: The unfounded claim that social media companies censor conservatives (tech. rep.). New York University Stern Center for Business and Human Rights

Barry D (2017). In a swirl of 'untruths' and 'falsehoods,' calling a lie a lie. New York Times. https://www.nytimes.com/2017/01/25/business/media/donald-trump-lie-media.html

Barsever D, Singh S, Neftci E (2020) Building a better lie detector with BERT: The difference between truth and lies. 2020 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn48605.2020.9206937

Bean SJ (2011) Emerging and continuing trends in vaccine opposition website content. Vaccine 29:1874–1880. https://doi.org/10.1016/j.vaccine.2011.01.003

Benkler Y, Faris R, Roberts H (2018) Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press

Berk R, Heidari H, Jabbari S, Kearns M, Roth A (2021) Fairness in criminal justice risk assessments: The state of the art. Sociol Methods Res 50:3–44. https://doi.org/10.1177/0049124118782533

Berlinski N, Doyle M, Guess AM, Levy G, Lyons B, Montgomery JM, Nyhan B, Reifler J (2021) The effects of unsubstantiated claims of voter fraud on confidence in elections. J Exp Polit Sci 10(1), 34–49. https://doi.org/10.1017/xps.2021.18

Bernstein A (2023) Republican Rep. Jim Jordan issues sweeping information requests to universities researching disinformation. Pro Publica. https://www.propublica.org/article/jim-jordan-disinformation-subpoena-universities

Birks J (2019) Fact-checking journalism and political argumentation: A British perspective. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-30573-4
Blair RA, Gottlieb J, Nyhan B, Paler L, Argote P, Stainfield CJ (2024) Interventions to counter misinformation: Lessons from the Global North and applications to the Global South. Curr Opin Psychol 55:101732. https://doi.org/10.1016/j.copsyc.2023.101732
Blake A (2023) Jenna Ellis's tearful guilty plea should worry Rudy Giuliani. Washington Post. https://www.washingtonpost.com/politics/2023/10/24/jenna-ellis-guilty-plea-georgia-giuliani-trump/
Blassnig S, Büchel F, Ernst N, Engesser S (2019) Populism and informal fallacies: an analysis of right-wing populist rhetoric in election campaigns. Argumentation 33:107–136. https://doi.org/10.1007/s10503-018-9461-2
Blitzer J (2023) Jim Jordan's conspiratorial quest for power. The New Yorker. https://www.newyorker.com/magazine/2023/10/30/jim-jordans-conspiratorial-quest-for-power
Bolin JL, Hamilton LC (2018) The news you choose: News media preferences amplify views on climate change. Environ Polit. https://doi.org/10.1080/09644016.2018.1423909
Bond CF, DePaulo BM (2006) Accuracy of deception judgments. Personal Soc Psychol Rev 10:214–234. https://doi.org/10.1207/s15327957pspr1003_2
Borchers C (2016) Why the New York Times decided it is now okay to call Donald Trump a liar. Washington Post. https://www.washingtonpost.com/news/the-fix/wp/2016/09/22/why-the-new-york-times-decided-it-is-now-okay-to-call-donald-trump-a-liar/
Borchers C (2017) Alex Jones should not be taken seriously, according to Alex Jones's lawyers. Washington Post. https://www.washingtonpost.com/news/the-fix/wp/2017/04/17/trump-called-alex-jones-amazing-joness-own-lawyer-calls-him-a-performance-artist/
Braun MT, Swol LMV, Vang L (2015) His lips are moving: Pinocchio effect and other lexical indicators of political deceptions. Discourse Process 52:1–20. https://doi.org/10.1080/0163853X.2014.942833
Brown TR (2022) Demystifying mindreading for the law. Wisconsin Law Review Forward, 1–11
Brulle RJ (2018) The climate lobby: A sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016. Climatic Change. https://doi.org/10.1007/s10584-018-2241-z
Brumback K (2023) Giuliani concedes he made public comments falsely claiming Georgia election workers committed fraud. Associated Press. https://apnews.com/article/giuliani-georgia-election-workers-lawsuit-false-statements-afc64a565ee778c6914a1a69dc756064
Bursztyn L, Egorov G, Enikolopov R, Petrova M (2019) Social media and xenophobia: Evidence from Russia (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w26567
Bursztyn L, Rao A, Roth C, Yanagizawa-Drott D (2020) Misinformation during a pandemic (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w27417
Canon DT, Sherman O (2021) Debunking the "Big Lie": Election Administration in the 2020 Presidential Election. Pres Stud Q 51:546–581. https://doi.org/10.1111/psq.12721
Carrasco-Farré C (2022) The fingerprints of misinformation: How deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. Hum Soc Sci Commun 9:1–18. https://doi.org/10.1057/s41599-022-01174-9
Cataldo JK, Bero LA, Malone RE (2010) "A delicate diplomatic situation": Tobacco industry efforts to gain control of the Framingham study. J Clin Epidemiol 63:841–853. https://doi.org/10.1016/j.jclinepi.2010.01.021
Ceccarelli L (2011) Manufactured scientific controversy: Science, rhetoric, and public debate. Rhetor Public Aff 14:195–228
Constâncio AS, Tsunoda DF, Silva HDFN, Silveira JMD, Carvalho DR (2023) Deception detection with machine learning: a systematic review and statistical analysis. PLoS One 18:e0281323. https://doi.org/10.1371/journal.pone.0281323
Cook J, Ellerton P, Kinkead D (2018) Deconstructing climate misinformation to identify reasoning errors. Environ Res Lett 13:024018
Cooley A, Nexon DH (2022) The real crisis of global order: Illiberalism on the rise. Foreign Aff 101:103–118
Darcy O (2021) Fox has quietly implemented its own version of a vaccine passport while its top personalities attack them. CNN. https://edition.cnn.com/2021/07/19/media/fox-vaccine-passport/index.html
Davis D, Sinnreich A (2020) Beyond fact-checking: Lexical patterns as lie detectors in Donald Trump's tweets. Int J Commun 14:5237–5260
de Freitas Melo P, Vieira CC, Garimella K, de Melo POV, Benevenuto F (2019) Can WhatsApp counter misinformation by limiting message forwarding? International Conference on Complex Networks and Their Applications, 372–384. https://doi.org/10.1007/978-3-030-36687-2_31
DellaVigna S, Kaplan E (2007) The fox news effect: media bias and voting. Q J Econ 122:1187–1234
Desikan A, MacKinney T, Kalman C, Carter JM, Reed G, Goldman GT (2023) An equity and environmental justice assessment of anti-science actions during the Trump administration. J Public Health Policy 44:147–162. https://doi.org/10.1057/s41271-022-00390-6
Dieterich W, Mendoza C, Brennan T (2016) COMPAS risk scales: Demonstrating accuracy equity and predictive parity. (tech. rep.). Northpoint, Inc
Dixit P, Mac R (2018) How WhatsApp destroyed a village. BuzzFeed News. https://www.buzzfeednews.com/article/pranavdixit/whatsapp-destroyed-village-lynchings-rainpada-india
Douglis A (2018) Disentangling perjury and lying. Yale J Law Hum 29:339–374
Dourado T, Salgado S (2021) Disinformation in the Brazilian pre-election context: Probing the content, spread and implications of fake news about Lula da Silva. Commun Rev 24:297–319. https://doi.org/10.1080/10714421.2021.1981705
Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv 4:eaao5580
Ecker UKH, Lewandowsky S, Apai J (2011) Terrorists brought down the plane!—No, actually it was a technical fault: Processing corrections of emotive information. Q J Exp Psychol 64:283–310. https://doi.org/10.1080/17470218.2010.497927
Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, Kendeou P, Vraga EK, Amazeen MA (2022) The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol 1:13–29. https://doi.org/10.1038/s44159-021-00006-y
Eggers AC, Garro H, Grimmer J (2021) No evidence for systematic voter fraud: a guide to statistical claims about the 2020 election. Proc Natl Acad Sci USA 118:e2103619118. https://doi.org/10.1073/pnas.2103619118
Enders A, Farhart C, Miller J, Uscinski J, Saunders K, Drochon H (2022) Are republicans and conservatives more likely to believe conspiracy theories? Polit Behav 1–24. https://doi.org/10.1007/s11109-022-09812-3
Enders AM, Uscinski JE (2021) Are misinformation, antiscientific claims, and conspiracy theories for political extremists? Group Processes & Intergroup Relations
Fallin A, Grana R, Glantz SA (2013) 'to quarterback behind the scenes, third-party efforts': the tobacco industry and the tea party. Tob Control 0:1–10. https://doi.org/10.1136/tobaccocontrol-2012-050815
Farber HJ, Neptune ER, Ewart GW (2018) Corrective statements from the tobacco industry: more evidence for why we need effective tobacco control. Ann Am Thorac Soc 15:127–130. https://doi.org/10.1513/annalsats.201711-845gh
Farrell H, Schneier B (2018) Common-knowledge attacks on democracy (tech. rep.). Berkman Klein Center for Internet & Society
Farrell J (2016) Network structure and influence of the climate change counter-movement. Nat Clim Change 6:370–374. https://doi.org/10.1038/nclimate2875
Fausset R, Hakim D (2023) Sidney Powell pleads guilty in Georgia Trump case. New York Times. https://www.nytimes.com/2023/10/19/us/sidney-powell-guilty-plea-trump-georgia.html
Fazio LK, Brashier NM, Payne BK, Marsh EJ (2015) Knowledge does not protect against illusory truth. J Exp Psychol General. https://doi.org/10.1037/xge0000098
Fazio L (2020) Pausing to consider why a headline is true or false can help reduce the sharing of false news. Harv Kennedy School Misinform Rev. https://doi.org/10.37016/mr-2020-009
Feldman L, Maibach EW, Roser-Renouf C, Leiserowitz A (2012) Climate on cable: the nature and impact of global warming coverage on Fox News, CNN, and MSNBC. Int J Press/Polit 17:3–31
Field H, Vanian J (2023) Tech layoffs ravage the teams that fight online misinformation and hate speech. CNBC. https://www.cnbc.com/2023/05/26/tech-companies-are-laying-off-their-ethics-and-safety-teams-.html
Fong A, Roozenbeek J, Goldwert D, Rathje S, van der Linden S (2021) The language of conspiracy: a psychological analysis of speech used by conspiracy theorists and their followers on Twitter. Group Process Intergroup Relat 24:606–623. https://doi.org/10.1177/1368430220987596
Francey N, Chapman S (2000) "operation Berkshire": the international tobacco companies' conspiracy. Br Med J 321:371–374. https://doi.org/10.1136/bmj.321.7257.371
Garrett RK, Bond RM (2021) Conservatives' susceptibility to political misperceptions. Sci Adv 7(23):eabf1234. https://doi.org/10.1126/sciadv.abf1234
Ghanem B, Rosso P, Rangel F (2020) An emotional analysis of false information in social media and news articles. ACM Trans Internet Technol 20:19:1–19:18. https://doi.org/10.1145/3381750
Goldberg B (2023) Defanging disinformation's threat to Ukrainian refugees. Jigsaw. https://medium.com/jigsaw/defanging-disinformations-threat-to-ukrainian-refugees-b164dbbc1c60
González-Bailón S, Lazer D, Barberá P, Zhang M, Allcott H, Brown T, Crespo-Tenorio A, Freelon D, Gentzkow M, Guess AM, Iyengar S, Kim YM, Malhotra N, Moehler D, Nyhan B, Pan J, Rivera CV, Settle J, Thorson E, Tucker

JA (2023) Asymmetric ideological segregation in exposure to political news on Facebook. Science 381:392–398. https://doi.org/10.1126/science.ade7138
Graham MH, Yair O (2023) Expressive responding and trump's big lie. Polit Behav. https://doi.org/10.1007/s11109-023-09875-w
Greene KT (2024) Partisan differences in the sharing of low-quality news sources by U.S. political elites. Polit Commun 1–20. https://doi.org/10.1080/10584609.2024.2306214
Grice HP (1975) Logic and conversation. In P Cole & JL Morgan (Eds.), Syntax and semantics, vol. 3: Speech acts (pp. 41–58). Academic Press
Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. Science 363:374–378. https://doi.org/10.1126/science.aau2706
Grofman B, Cervas J (2023) Statistical fallacies in claims about 'massive and widespread fraud' in the 2020 presidential election: examining claims based on aggregate election results 1,2. Stat Public Policy 1–36. https://doi.org/10.1080/2330443X.2023.2289529
Guess AM, Nyhan B, Reifler J (2020a) Exposure to untrustworthy websites in the 2016 U.S. election. Nat Hum Behav 4:472–480. https://doi.org/10.1038/s41562-020-0833-x
Guess AM, Lockett D, Lyons B, Montgomery JM, Nyhan B, Reifler J (2020b) "Fake news" may have limited effects on political participation beyond increasing beliefs in false claims. Harv Kennedy School Misinform Rev 1(1). https://doi.org/10.37016/mr-2020-004
Guess AM, Nagler J, Tucker J (2019) Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv 5:eaau4586. https://doi.org/10.1126/sciadv.aau4586
Harris KR (2022) Real fakes: The epistemology of online misinformation. Philos Technol 35. https://doi.org/10.1007/s13347-022-00581-9
Henricksen W, Betz B (2023) The stolen election lie and the freedom of speech. Penn State Law Review. https://doi.org/10.2139/ssrn.4354211
Hotez P (2023) Anti-science conspiracies pose new threats to US biomedicine in 2023. FASEB BioAdvances. https://doi.org/10.1096/fba.2023-00032
Hruschka TMJ, Appel M (2023) Learning about informal fallacies and the detection of fake news: An experimental intervention. PLoS One 18:e0283238
Hsu SS, Weiner R (2023) Defamed Georgia poll workers who won $148M from Giuliani sue him again. Washington Post. https://www.washingtonpost.com/dc-md-va/2023/12/18/giuliani-defamation-lawsuit-georgia/
Hurley L (2023) Supreme Court blocks restrictions on Biden administration efforts to get platforms to remove social media posts. NBC News. https://www.nbcnews.com/politics/supreme-court/supreme-court-blocks-biden-social-media-curbs-rcna105785
Huszár F, Ktena SI, O'Brien C, Belli L, Schlaikjer A, Hardt M (2022) Algorithmic amplification of politics on Twitter. Proc Natl Acad Sci 119:e2025334119. https://doi.org/10.1073/pnas.2025334119
Jacobson GC (2021) Donald Trump's big lie and the future of the republican party. Pres Stud Q 51:273–289. https://doi.org/10.1111/psq.12716
Jacobson GC (2023) The dimensions, origins, and consequences of belief in Donald Trump's Big Lie. Polit Sci Q 138:133–166. https://doi.org/10.1093/psquar/qqac030
Jalli N, Idris I (2019) Fake news and elections in two Southeast Asian nations: A comparative study of Malaysia general election 2018 and Indonesia presidential election 2019. Proceedings of the International Conference of Democratisation in Southeast Asia (ICDeSA 2019). https://doi.org/10.2991/icdesa-19.2019.30
Jung Y, Lee S (2023) Trump vs. the GOP: Political Determinants of COVID-19 Vaccination. Polit Behav. https://doi.org/10.1007/s11109-023-09882-x
Kellow CL, Steeves HL (1998) The role of radio in the Rwandan genocide. https://doi.org/10.1111/j.1460-2466.1998.tb02762.x
Kinser S (2020) Science in an age of scrutiny: How scientists can respond to criticism and personal attacks. Union of Concerned Scientists. https://www.ucsusa.org/sites/default/files/2020-09/science-in-an-age-of-scrutiny-2020.pdf
Kozyreva A, Herzog SM, Lewandowsky S, Hertwig R, Lorenz-Spreen P, Leiser M, Reifler J (2023) Resolving content moderation dilemmas between free speech and harmful misinformation. Proc Natl Acad Sci USA 120:e2210666120. https://doi.org/10.1073/pnas.2210666120
Kozyreva A, Lorenz-Spreen P, Herzog SM, Ecker UKH, Lewandowsky S, Hertwig R, Ali A, Bak-Coleman JB, Barzilai S, Basol M, Berinsky A, Betsch C, Cook J, Fazio LK, Geers M, Guess AM, Huang H, Larreguy H, Maertens R, … Wineburg S (2024) Toolbox of interventions against online misinformation. Nat Hum Behav. https://doi.org/10.31234/osf.io/x8ejt
Kozyreva A, Smillie L, Lewandowsky S (2023) Incorporating psychological science into policy making. Eur Psychol 28:206–224. https://doi.org/10.1027/1016-9040/a000493
Kozyreva A, Wineburg S, Lewandowsky S, Hertwig R (2023) Critical ignoring as a core competence for digital citizens. Curr Dir Psychol Sci 32:81–88. https://doi.org/10.1177/09637214221121570
Kuklinski JH, Quirk PJ, Schwieder DW, Rich RF (1998) "Just the facts, ma'am": political facts and public opinion. Ann Am Acad Political Soc Sci 560:143–154. https://doi.org/10.1177/0002716298560001011
Kull S, Ramsay C, Lewis E (2003) Misperceptions, the media, and the Iraq war. Political Sci Q 118:569–598
Kumari R, Ashok N, Ghosal T, Ekbal A (2022) *What the fake?* Probing misinformation detection standing on the shoulder of novelty and emotion. Inf Process Manag 59:102740. https://doi.org/10.1016/j.ipm.2021.102740
Lackey J (2013) Lies and deception: an unhappy divorce. Analysis. https://doi.org/10.1093/analys/ant006
Lagioia F, Rovatti R, Sartor G (2023) Algorithmic fairness through group parities? The case of COMPAS-SAPMOC. AI & SOCIETY, 38, 459–478. https://doi.org/10.1007/s00146-022-01441-y
Landman A, Glantz SA (2009) Tobacco industry efforts to undermine policy-relevant research. Am J Public Health 99:45–58. https://doi.org/10.2105/AJPH.2004.050963
Lasser J, Aroyehun ST, Simchon A, Carrella F, Garcia D, Lewandowsky S (2022) Social media sharing of low quality news sources by political elites. PNAS Nexus, pgac186. https://doi.org/10.1093/pnasnexus/pgac186
Latour B (2004) Why has critique run out of steam? From matters of fact to matters of concern. Crit Inq 30:225–248
Lebernegg N, Eberl J-M, Tolochko P, Boomgaarden H (2024) Do you speak disinformation? Computational detection of deceptive news-like content using linguistic and stylistic features. Digit J. https://doi.org/10.1080/21670811.2024.2305792
Leonhardt D (2021) Red Covid. New York Times. https://www.nytimes.com/2021/09/27/briefing/covid-red-states-vaccinations.html
Lerer L (2020) Giuliani in public: 'it's a fraud.' Giuliani in court: 'This is not a fraud case.' New York Times https://www.nytimes.com/2020/11/18/us/politics/trump-giuliani-voter-fraud.html
Levine S (2023) Angry Fox News chief said fact-checks of Trump's election lies 'bad for business'. The Guardian. https://www.theguardian.com/media/2023/mar/29/fox-news-trump-fact-check-election-lies-dominion
Lewandowsky S (2020) Willful construction of ignorance: A tale of two ontologies. In R Hertwig & C Engel (Eds.), Deliberate ignorance: Choosing not to know (pp. 101–117). MIT Press
Lewandowsky S, Ballard T, Oberauer K, Benestad R (2016) A blind expert test of contrarian claims about climate data. Glob Environ Change 39:91–97. https://doi.org/10.1016/j.gloenvcha.2016.04.013
Lewandowsky S, Ecker UKH, Cook J (2017) Beyond misinformation: understanding and coping with the post-truth era. J Appl Res Mem Cogn 6:353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
Lewandowsky S, Kalish ML, Ngang S (2002) Simplified learning in complex situations: Knowledge partitioning in function learning. J Exp Psychol Gen 131:163–193. https://doi.org/10.1037/0096-3445.131.2.163
Lewandowsky S, Robertson RE, DiResta R (2023a) Challenges in understanding human-algorithm entanglement during online information consumption. Perspect Psychol Sci. https://doi.org/10.1177/17456916231180809
Lewandowsky S, Stritzke WGK, Freund AM, Oberauer K, Krueger JI (2013) Misinformation, disinformation, and violent conflict: From Iraq and the "War on Terror" to future threats to peace. Am Psychol 68:487–501. https://doi.org/10.1037/a0034515
Lewandowsky S (2022) Fake news and participatory propaganda. In R Pohl (Ed.), Cogn illusions (pp. 324–340). Routledge https://doi.org/10.4324/9781003154730-23
Lewandowsky S, Ecker UKH, Cook J, van der Linden S, Roozenbeek J, Oreskes N (2023b) Misinformation and the epistemic integrity of democracy. Curr Opin Psychol 101711. https://doi.org/10.1016/j.copsyc.2023.101711
Lewandowsky S, Pomerantsev P (2022) Technology and democracy: a paradox wrapped in a contradiction inside an irony. Memory Mind Media 1. https://doi.org/10.1017/mem.2021.7
Lewandowsky S, van der Linden S (2021) Countering misinformation and fake news through inoculation and prebunking. Eur Rev Soc Psychol 32:348–384. https://doi.org/10.1080/10463283.2021.1876983
Li D (2004) Echoes of violence: considerations on radio and genocide in Rwanda. J Genocide Res 6:9–27. https://doi.org/10.1080/1462352042000194683
Lin H, Lasser J, Lewandowsky S, Cole R, Gully A, Rand DG, Pennycook G (2023) High level of correspondence across different news domain quality rating sets. PNAS Nexus 2:pgad286. https://doi.org/10.1093/pnasnexus/pgad286
Lorenz-Spreen P, Oswald L, Lewandowsky S, Hertwig R (2022) A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nat Hum Behav 1–28. https://doi.org/10.1038/s41562-022-01460-1
Lu C, Hu B, Li Q, Bi C, Ju X-D (2023) Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: systematic review and meta-analysis. J Med Internet Res 25:e49255. https://doi.org/10.2196/49255

Martel C, Allen J, Pennycook G, Rand DG (2024) Crowds can effectively identify misinformation at scale. Perspect Psychol Sci 19:477–488. https://doi.org/10.1177/17456916231190388

Martel C, Pennycook G, Rand DG (2020) Reliance on emotion promotes belief in fake news. Cogn Res Princ Implic 5:47. https://doi.org/10.1186/s41235-020-00252-3

Mattes K, Popova V, Evans JR (2023) Deception detection in politics: can voters tell when politicians are lying. Polit Behav 45:395–418. https://doi.org/10.1007/s11109-021-09747-1

McGraw KM (1998) Manipulating public opinion with moral justification. Ann Am Acad Political Soc Sci 560:129–142. https://doi.org/10.1177/0002716298560001010

McIntyre L (2018) Post-truth. MIT Press

McLauchlin T (2023) Tail risks for 2024: Prospects for a violent constitutional crisis in the United States (tech. rep. No. 28). Network for Strategic Analysis, Queen's University, Canada

Mounk Y (2023) The identity trap. Penguin Random House

Müller K, Schwarz C (2021) Fanning the flames of hate: social media and hate crime. J Eur Econ Assoc 19:2131–2167. https://doi.org/10.1093/jeea/jvaa045

Musi E, Aloumpi M, Carmi E, Yates S, O'Halloran K (2022) Developing fake news immunity: Fallacies as misinformation triggers during the pandemic. Online J Commun Media Technol 12:e202217. https://doi.org/10.30935/ojcmt/12083

Musi E, Reed C (2022) From fallacies to semi-fake news: improving the identification of misinformation triggers across digital media. Discourse Soc 33:349–370. https://doi.org/10.1177/09579265221076609

Muzaffar M (2021) Tucker Carlson admits he lies on his show: 'I really try not to… [but] I certainly do'. The Independent. https://www.independent.co.uk/news/world/americas/tucker-carlson-fox-news-dave-rubin-b1919738.html

Nadarevic L, Reber R, Helmecke AJ, Köse D (2020) Perceived truth of statements and simulated social media postings: An experimental investigation of source credibility, repeated exposure, and presentation format. Cogn Res Princ Implic 5. https://doi.org/10.1186/s41235-020-00251-4

Nan X, Wang Y, Thier K (2022) Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. Soc Sci Med 314:115398. https://doi.org/10.1016/j.socscimed.2022.115398

Neff A, Fredrickson C (2023) Trump's lawyers face sanctions, discipline, and indictment – how should the legal profession respond? Just Security. https://www.justsecurity.org/90509/trumps-lawyers-face-sanctions-discipline-and-indictment-how-should-the-legal-profession-respond/

Neo R (2022) A cudgel of repression: analysing state instrumentalisation of the 'fake news' label in Southeast Asia. Journalism 23:1919–1938. https://doi.org/10.1177/1464884920984060

Nieminen S, Sankari V (2021) Checking PolitiFact's fact-checks. J Stud 22:358–378. https://doi.org/10.1080/1461670x.2021.1873818

Nix N, Menn J (2023) These academics studied falsehoods spread by Trump. Now the GOP wants answers. Washington Post. https://www.washingtonpost.com/technology/2023/06/06/disinformation-researchers-congress-jim-jordan/

Nix N, Zakrzewski C, Menn J (2023) Misinformation research isbuckling under GOP legal attacks. Washington Post. https://www.washingtonpost.com/technology/2023/09/23/online-misinformation-jim-jordan/

Nyberg D (2023) The passive revolution is televised: The dominant ideology of media capitalism. Organization, 13505084231180288. https://doi.org/10.1177/13505084231180288

Ognyanova K, Lazer D, Robertson RE, Wilson C (2020) Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. Harv Kennedy School Misinform Rev. https://doi.org/10.37016/mr-2020-024

Oreskes N, Conway EM (2010) Merchants of doubt. Bloomsbury Publishing

Oreskes N, & Conway EM (2023) The big myth. New York City: Bloomsbury Publishing

O'Toole F (2022) We don't know ourselves: A personal history of Ireland since 1958. Head of Zeus

Painter DL, Fernandes J (2022) "the big lie": How fact checking influences support for insurrection. Am Behav Sci 000276422211031. https://doi.org/10.1177/00027642221103179

Papantoniou K, Papadakos P, Patkos T, Flouris G, Androutsopoulos I, Plexousakis D (2021) Deception detection in text and its relation to the cultural dimension of individualism/collectivism. Nat Lang Eng 28:545–606. https://doi.org/10.1017/s1351324921000152

Peltz M (2023) New details in Dominion suit reveal damning evidence of deception in Fox News' 2020 election coverage. Mediamatters. https://www.mediamatters.org/foxdominion-lawsuit/new-details-dominion-suit-reveal-damning-evidence-deception-fox-news-2020

Peng W, Lim S, Meng J (2023) Persuasive strategies in online health misinformation: a systematic review. Inf Commun Soc 26:2131–2148. https://doi.org/10.1080/1369118X.2022.2085615

Pennycook G, Rand DG (2021) Research note: Examining false beliefs about voter fraud in the wake of the 2020 presidential election. Harvard Kennedy School (HKS) Misinform Rev *2*. https://doi.org/10.37016/mr-2020-51

Pennycook G, Rand DG (2019) Fighting misinformation on social media using crowdsourced judgments of news source quality. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1806781116

Pereira FB, Bueno NS, Nunes F, Pavão N (2022) Fake news, fact checking, and partisanship: the resilience of rumors in the 2018 brazilian elections. J Polit 84:2188–2201. https://doi.org/10.1086/719419

Pereira FB, Bueno NS, Nunes F, Pavão N (2023) Inoculation reduces misinformation: experimental evidence from multidimensional interventions in brazil. J Exp Polit Sci 1–12. https://doi.org/10.1017/xps.2023.11

Persad G, Emanuel EJ, Sangenito S, Glickman A, Phillips S, Largent EA (2021) Public perspectives on COVID-19 vaccine prioritization. JAMA Netw Open 4:e217943. https://doi.org/10.1001/jamanetworkopen.2021.7943

Pillai RM, Fazio LK (2023) Explaining why headlines are true or false reduces intentions to share false information. Collabra: Psychol 9. https://doi.org/10.1525/collabra.87617

Pinna M, Picard L, Goessmann C (2022) Cable news and COVID-19 vaccine uptake. Sci Rep 12:16804. https://doi.org/10.1038/s41598-022-20350-0

Polantz K (2021) Lawyers sanctioned for 'conspiracy theory' election fraud lawsuit. CNN. https://edition.cnn.com/2021/08/04/politics/lawyers-colorado-2020-election/index.html

Porter E, Wood TJ (2021) The global effectiveness of fact-checking: evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. Proc Natl Acad Sci USA 118:e2104235118. https://doi.org/10.1073/pnas.2104235118

Pothos EM, Lewandowsky S, Basieva I, Barque-Duran A, Tapper K, Khrennikov A (2021) Information overload for (bounded) rational agents. Proc R Soc B Biol Sci 288:20202957. https://doi.org/10.1098/rspb.2020.2957

Prike T, Butler LH, Ecker UKH (2024) Source-credibility information and social norms improve truth discernment and reduce engagement with misinformation online. Sci Rep 14:6900. https://doi.org/10.1038/s41598-024-57560-7

Proctor RN (2011) Golden holocaust: Origins of the cigarette catastrophe and the case for abolition. University of California Press

Proctor RN (2012) The history of the discovery of the cigarette–lung cancer link: evidentiary traditions, corporate denial, global toll. Tob Control 21(2):87–91. https://doi.org/10.1136/tobaccocontrol-2011-050338

Pröllochs N, Bär D, Feuerriegel S (2021) Emotions explain differences in the diffusion of true vs. false social media rumors. Sci Rep 11:22721. https://doi.org/10.1038/s41598-021-01813-2

Reid T (2022) Voting machine problems in Arizona seized on by Trump, election deniers. Reuters. https://www.reuters.com/world/us/voting-machine-problems-battleground-arizona-seized-by-trump-election-deniers-2022-11-08/

Rogers T, Zeckhauser R, Gino F, Norton MI, Schweitzer ME (2017) Artful paltering: the risks and rewards of using truthful statements to mislead others. J Personal Soc Psychol 112:456–473

Roozenbeek J, van der Linden S, Goldberg B, Rathje S, Lewandowsky S (2022) Psychological inoculation improves resilience against misinformation on social media. Sci Adv 8:eabo6254. https://doi.org/10.1126/sciadv.abo6254

Roozenbeek J, Suiter J, Culloty E (2023) Countering misinformation: evidence, knowledge gaps, and implications of current interventions. Eur Psychol. https://doi.org/10.31234/osf.io/b52um

Rumpler E, Feldman JM, Bassett MT, Lipsitch M (2023) Fairness and efficiency considerations in COVID-19 vaccine allocation strategies: A case study comparing front-line workers and 65–74 year olds in the United States. PLOS Glob Public Health 3:e0001378. https://doi.org/10.1371/journal.pgph.0001378

Rutenberg J, Myers SL (2024) How Trump's allies are winning the war over disinformation. New York Times. https://www.nytimes.com/2024/03/17/us/politics/trump-disinformation-2024-social-media.html

Salek TA (2023) Deflecting deliberation through rhetorical nihilism: "Stop the Steal" as an unethical and intransigent rival public. Commun Democracy 57:94–118. https://doi.org/10.1080/27671127.2023.2202744

Scarcella M (2023) US Supreme Court rebuffs lawyers punished after 'woeful' suit backing Trump. Reuters. https://www.reuters.com/legal/us-supreme-court-rebuffs-lawyers-punished-after-woeful-suit-backing-trump-2023-10-02/

Shahid F, Vashistha A (2023) Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony? Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3544548.3581538

Simonov A, Sacher S, Dubé J-P, Biswas S (2020) The persuasive effect of Fox News: Non-compliance with social distancing during the Covid-19 pandemic (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w27237

Simonov A, Sacher S, Dubé J-P, Biswas S (2022) Frontiers: the persuasive effect of fox news: noncompliance with social distancing during the COVID-19 pandemic. Mark Sci 41:230–242. https://doi.org/10.1287/mksc.2021.1328

Smith P, Bansal-Travers M, O'Connor R, Brown A, Banthin C, Guardino-Colket S, Cummings K (2011) Correcting over 50 years of tobacco industry misinformation. Am J Prev Med 40:690–698

Soroka S, Fournier P, Nir L (2019) Cross-national evidence of a negativity bias in psychophysiological reactions to news. Proc Natl Acad Sci USA 116:18888–18892. https://doi.org/10.1073/pnas.1908369116

Stapleton A (2016) No, you can't vote by text message. CNN. https://edition.cnn.com/2016/11/07/politics/vote-by-text-message-fake-news/index.html

Starbird K, DiResta R, DeButts M (2023) Influence and Improvisation: participatory disinformation during the 2020 US election. Soc Media Soc 9:20563051231177943. https://doi.org/10.1177/20563051231177943

Supran G, Rahmstorf S, Oreskes N (2023) Assessing ExxonMobil's global warming projections. Science 379(6628):eabk0063. https://doi.org/10.1126/science.abk0063

Supran G, Oreskes N (2017) Assessing ExxonMobil's climate change communications (1977–2014). Environ Res Lett 12:084019. https://doi.org/10.1088/1748-9326/aa815f

Supran G, Oreskes N (2021) Rhetoric and frame analysis of ExxonMobil's climate change communications. One Earth. https://doi.org/10.1016/j.oneear.2021.04.014

Swire-Thompson B, Ecker UKH, Lewandowsky S, Berinsky AJ (2020) They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. Polit Psychol 41:21–34. https://doi.org/10.1111/pops.12586

Takhshid Z (2021) Regulating social media in the global south. Vanderbilt J Entertain Technol Law 24:1–56

Tenove C (2020) Protecting democracy from disinformation: Normative threats and policy responses. Int J Press/Polit 25:517–537. https://doi.org/10.1177/1940161220918740

Terkel A, Timm JC, Gregorian D (2023) Here's what fox news was trying to hide in its dominion lawsuit redactions. NBC News. https://www.nbcnews.com/politics/elections/dominion-releases-previously-redacted-slides-fox-news-lawsuit-rcna77257

U.S. House of Representatives Judiciary Committee. (2023). News release: Jim Jordan on why the select subcommittee on the weaponization of the federal government is necessary | House Judiciary Committee Republicans. http://judiciary.house.gov/media/press-releases/jim-jordan-on-why-the-select-subcommittee-on-the-weaponization-of-the-federal

Uscinski JE, Parent JM (2014) American conspiracy theories. Oxford University Press

Uscinski JE (2015) The epistemology of fact checking (is still naìve): Rejoinder to Amazeen. Crit Rev 27:243–252. https://doi.org/10.1080/08913811.2015.1055892

Van Der Zee S, Poppe R, Havrileck A, Baillon A (2021) A personal model of trumpery: linguistic deception detection in a real-world high-stakes setting. Psychol Sci. https://doi.org/10.1177/09567976211015941

van Doorn M (2023) Advancing the debate on the consequences of misinformation: clarifying why it's not (just) about false beliefs. Inquiry 0:1–27. https://doi.org/10.1080/0020174X.2023.2289137

Vese D (2022) Governing fake news: the regulation of social media and the right to freedom of expression in the era of emergency. Eur J Risk Regul 13:477–513. https://doi.org/10.1017/err.2021.48

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359:1146–1151. https://doi.org/10.1126/science.aap9559

Walasek L, Brown GDA (2023) Incomparability and incommensurability in choice: No common currency of value? Perspect Psychol Sci. https://doi.org/10.1177/17456916231192828

Wallace J, Goldsmith-Pinkham P, Schwartz JL (2023) Excess death rates for republican and democratic registered voters in Florida and Ohio during the COVID-19 pandemic. JAMA Internal Med. https://doi.org/10.1001/jamainternmed.2023.1154

Wanless A, Berk M (2019) The audience is the amplifier: Participatory propaganda. In P Baines, N O'Shaughnessy, & N Snow (Eds.), The sage handbook of propaganda (pp. 85–104). Sage

Wardle C, Derakhshan H (2017) Information disorder: Toward an interdisciplinary framework for research and policymaking (tech. rep.). Council of Europe. https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77

Weinschenk AC, Panagopoulos C, van der Linden S (2021) Democratic norms, social projection, and false consensus in the 2020 U.S. presidential election. J Polit Mark 20:255–268. https://doi.org/10.1080/15377857.2021.1939568

West D (2023) We shouldn't turn disinformation into a constitutional right. Brookings Institution. https://www.brookings.edu/articles/we-shouldnt-turn-disinformation-into-a-constitutional-right/

Williams D (2021) Motivated ignorance, rationality, and democratic politics. Synthese 198:7807–7827. https://doi.org/10.1007/s11229-020-02549-8

Williams D (2022) The marketplace of rationalizations. Econ Philosophy, 1–25. https://doi.org/10.1017/S0266267121000389

Wu T (2017) The attention merchants. Atlantic Books

Yee AK (2023a) Information deprivation and democratic engagement. Philos Sci 90:1110–1119. https://doi.org/10.1017/psa.2023.9

Yee AK (2023b) Machine learning, misinformation, and citizen science. Eur J Philosophy Sci 13. https://doi.org/10.1007/s13194-023-00558-1

Zakrzewski C, Lima C, Harwell D (2023) What the Jan. 6 probe found out about social media, but didn't report. Washington Post. https://www.washingtonpost.com/technology/2023/01/17/jan6-committee-report-social-media/

Acknowledgements

SL acknowledges financial support from the European Research Council (ERC Advanced Grant 101020961 PRODEMINFO), the Humboldt Foundation through a research award, the Volkswagen Foundation (grant "Reclaiming individual autonomy and democratic discourse online: How to rebalance human and algorithmic decision making"), and the European Commission (Horizon 2020 grants 964728 JITSUVAX and 101094752 SoMe4Dem). SL also receives funding from Jigsaw (a technology incubator created by Google) and from UK Research and Innovation through EU Horizon replacement funding grant number 10049415. UKHE acknowledges support from the Australian Research Council (grant FT190100708). For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

The first author created the first draft and all other authors contributed additional material and comments and suggestions and participated jointly in the editing and revision process.

Competing interests

SL, JR, and SvdL have received funding from Google Jigsaw for empirical work on inoculation against misinformation and continue to collaborate with Jigsaw. NO has received funding from the Rockefeller Family Fund to support research on fossil fuel industry disinformation. She has also served as a consultant to the law firm Sher-Edling, who are representing several counties in California suing the fossil fuel industry, and as an expert witness in the defamation case of climate scientist Michael Mann. The remaining authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors that would require consent.

Additional information

Correspondence and requests for materials should be addressed to Stephan Lewandowsky.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024

Document 9 - Transparency International UK

Y Pwyllgor Safonau Ymddygiad

Standards of Conduct Committee

By email

5 December 2024

Individual Member Accountability Inquiry

Dear Hannah,

Thank you for your letter and your request for information about TI-UK's position in relation to a duty of candour. This note seeks to provide a response to your letter as well as some supplementary information we hope the Committee finds helpful.

1. The Public Authority (Accountability) Bill and a duty of candour

As you will be aware, the Public Authority (Accountability) Bill as proposed by Andy Burnham MP fell when the 2017 election was called. Since then, the new Labour government has committed to introducing a duty of candour as part of a so called 'Hillsborough Law' as included in the King's Speech.

At TI-UK we would be minded to support any legislation which improved the functioning of public inquiries as a vital tool for increased transparency and accountability. However, we do not yet have a finalised position on any proposed legislation from the current government in this area as no drafts have yet been published. As we made clear in our evidence to the Committee, our recommendations for reform are evidence based and pragmatic. We spend time considering how objectives will be reached, and whether proposed levers will accomplish the stated aims.

We can see the potential for parallels between a duty of candour and an offence of deception, especially around the importance of definition but we would not want to make assumptions until there is more clarity on the proposal at hand.

We note that JUSTICE were a lead organisation involved in developing Mr Burnham's proposed legislation and would suggest they may be better positioned to answer your queries.

2. TI-UK's position in relation to amendments proposing suspension of parliamentary privilege in the Bribery Act / Parliamentary Standards Act

We were asked a specific question by Adam Price MS about the 2010 Bribery Act and the Parliamentary Standards Act 2009. Mr Price wanted to know if Transparency International UK took a position on the removal of parliamentary privilege in relation to acts of bribery.

I have reviewed our files and cannot find any reference to this particular aspect of legislative reform in relation to either the Bribery Act or the Parliamentary Standards Act.

We did express significant concerns with the Parliamentary Standards Act as first introduced. We have consistently presented a number of recommendations for reform, many of which have been adopted and have improved the system at Westminster. As we said in our oral evidence, we propose

a package of reforms alongside strengthening the standards regimes which together would bring enhanced transparency and accountability.

As mentioned in our oral evidence, TI-UK has some long-standing recommendations to improve transparency and accountability, which we believe would enhance public trust in politics. These include:

- Lobbying transparency.
- Caps on political donations and spending.
- Improved reporting of gifts and hospitality.
- Better regulation of the revolving door between political life and activity where your previous political role might bring undue influence.
- Improving ministerial accountability.
- Strengthening accountability mechanisms like standards regimes.

All of these proposals are intended to increase transparency – we are unconvinced that a court led accountability process would provide this increased transparency. The courts may seem transparent to those who encounter them regularly and understand their ways of working but this is not necessarily true of the general public. The aforementioned JUSTICE also work to increase public understanding of the judicial system.

3. Parliamentary Privilege

In relation to parliamentary privilege, the committee may find this resolution from the Parliamentary Assembly of the Council of Europe (PACE) of interest. It states that parliamentary immunity exists to guarantee the independence of elected representatives in order for them to exercise their democratic functions effectively, and to protect the parliamentary institution (PACE 2016). An MP's freedom of speech and of expression cannot be constrained since it is essential to the performance of their institutional role.

PACE goes on to add a caveat which may be of interest to the committee:
"The absolute protection of the acts and statements of members of parliament, especially as far as hate speech is concerned, does indeed pose a problem in view of the current rise in extremism and nationalism against the backdrop of an upsurge in terrorism and the migration crisis. The Assembly notes and welcomes the fact that, in some States, insulting or defamatory utterances, incitement to hatred or violence or, in particular, racist remarks are not covered by non-liability rules."

It is worth considering that parliamentary privilege may feel like a two-tier system but in countries where democracy and democratic institutions are more fragile, the principle of parliamentary immunity is a vital protection for opposition politicians.

That said, impunity is not immunity, and offences which are unrelated to the politician's role, or which constitute corruption should be treated appropriately.

4. International examples

We also wanted to take this opportunity to share some international examples with the committee.

Firstly, this Transparency International Anti-Corruption Helpdesk briefing on codes of conduct for parliamentarians includes some universal principles as well as some different approaches.
https://knowledgehub.transparency.org/assets/uploads/kproducts/Code-of-Conduct-for-Parliamentarians-2022.pdf

Secondly, you may also be interested in Parliamentary Integrity: A Resource for Reformers, a report written for the Organization for Security and Co-operation in Europe. As the introductory webpage describes: "T*he aim of this study is to identify the main concerns and possible obstacles that need to be considered while reforming, developing and designing parliamentary standards of conduct, including, but not limited to, codes of conduct*." The author, Professor Elizabeth David-Barrett sits on the TI-UK board.

Thirdly, in relation to how deception causes democratic harm: When it comes to combatting disinformation and thinking about how to counter the impact of deception and disinformation on democratic discourse, there are many excellent examples of different levels of activity being delivered globally by member countries of the Open Government Partnership. Using open government principles of transparency, accountability, integrity and participation could protect democratic freedoms both online and off.

Some interesting approaches include in The Netherlands where the government committed to introducing greater transparency into how political parties are funded while making online election campaigns and political advertisements more transparent, as a means of combatting disinformation.

Then in Uruguay, six of their political parties signed an Ethical Pact Against Disinformation that pledged "not to generate or promote false news or disinformation campaigns to the detriment of political adversaries."

And France committed to hosting multi-stakeholder dialogues with civil society and research institutions, to identify research priorities and existing tools, resources, and techniques to monitor and counter misinformation and disinformation. The government also committed to discussing proposed solutions to counter the dissemination of misinformation and disinformation.

More information is available here https://www.opengovpartnership.org/stories/addressing-harmful-information-online/

I hope the above is of assistance as the Committee continues its inquiry. Do get in touch with any further questions.

Yours

Juliet Swann

10 Queen Street Place, London, EC4R 1BE | Email: juliet.swann@transparency.org.uk | Tel: + 44 (0)7968 527561

Huw Irranca-Davies AS/MS
Y Dirprwy Brif Weinidog ac Ysgrifennydd y Cabinet dros Newid Hinsawdd a Materion Gwledig
Deputy First Minister and Cabinet Secretary for Climate Change and Rural Affairs

Ein cyf/Our ref: HIDCC/PO/0398/24

Hannah Blythyn MS
Chair
Standards of Conduct Committee
Welsh Parliament
Cardiff Bay
Cardiff
CF99 1SN

2 January 2025

Dear Hannah,

Thank you for your letter, and for the opportunity to provide evidence to the Standards of Conduct Committee in respect of your inquiry into individual member accountability. I will respond to your requests for additional information in turn.

What preparatory work the Government has undertaken following the commitment made in the Senedd.

Following the Government's commitment to invite the Committee to make proposals in respect of deceptive statements, my officials have been working through the policy and legal implications of the three options identified by the Committee in the consultation – a criminal offence brought before the criminal courts, a civil offence adjudicated through an independent judicial process, or strengthening the existing Code of Conduct and standards procedures.

Officials have been working through the detailed issues and have been able to determine the initial scope of policy decisions that will be required. I will therefore be in a position to begin taking those decisions once the Committee's recommendations are known and can appropriately be taken into account.

Consideration of the procedure and processes for giving effect to its commitment and assessing any resource or capacity implications.

As is the case in the development of all Government legislation, policy development is being undertaken alongside detailed analysis from Legal Services, in order to be ready to instruct our legislative drafters. Elements of the policy and legal capacity that was previously deployed in the development and delivery of the Senedd Cymru (Members and Elections) Act have been retained for this purpose.

Canolfan Cyswllt Cyntaf / First Point of Contact Centre:
0300 0604400
Bae Caerdydd • Cardiff Bay
Caerdydd • Cardiff
CF99 1SN
Gohebiaeth.Huw.Irranca-Davies@llyw.cymru
Correspondence.Huw.Irranca-Davies@gov.wales

Rydym yn croesawu derbyn gohebiaeth yn Gymraeg. Byddwn yn ateb gohebiaeth a dderbynnir yn Gymraeg yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi.

We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

In order to meet our commitments, we are seeking to make progress at pace. This may mean that decisions will need to be taken as to the appropriate balance between what is included within primary legislation, and what can be provided for in secondary legislation. In addition, the constitutional significance of any legislative proposals in this space means that it is essential that they receive the appropriate scrutiny and consideration from the Senedd.

What timeframe you consider would be necessary to develop and bring through legislation of this sort. Perhaps you could assist the committee on this point with the length taken on any comparable legislation.

As I set out in my evidence to the Committee, we are preserving a slot for this legislation in the 5th year of our legislative programme.

The challenging timescales should not be understated however, and it is essential that all parts of the process are able to proceed at pace. A balance needs to be struck between maximising the time available to properly develop legislation in what is an area with limited precedent and ensuring that the Senedd has the necessary time to scrutinise and improve that legislation.

Any conversations you may have had with the UK Government about any potential impacts on the justice system and what if any response received.

I am conscious that this Committee's inquiry is currently ongoing, including a public consultation, and I have been mindful to not cut across or pre-empt the outcome of that engagement.

In respect of potential impact on the justice system, I am also aware that the Committee has taken evidence from the Chief Constable of North Wales Police and the Criminal Bar Association, as well as receiving written evidence from the [President] of the Adjudication Panel for Wales, and the Crown Prosecution Service. As the detailed policy positions are developed, and where appropriate, we will of course ensure that consideration of the views and expertise of key stakeholders are embedded in that process.

The Justice Impact Assessment process is utilised in the development of all legislation that impacts on the justice system, and that process involves engagement with the Ministry of Justice when the proposed policy model is sufficiently well defined. The assessment will form part of the Explanatory Memorandum that will accompany any Bill.

What your assessment would be of the risk of vexatious or malicious complaints if legislation is introduced.

It is important that any system strikes the right balance between providing for an effective model for member accountability, and not allowing the system to be used to prevent legitimate discourse, and I welcome the views of the Committee as to where that balance should lie. As part of the policy work set out above, my officials have been giving consideration as to how the parameters of the wrongdoing are drawn, as well as the extent to which a mechanism to filter complaints should form part of the system, to ensure only the most serious incidences of deception are captured. Additionally, were broader parliamentary privilege protections afforded to Members of the Senedd, the interactions between any legislation on deceptive statements and those protections would need to be carefully considered.

In conclusion, I fully agree with your assessment that *“legislating for deliberate deception has the potential to be complex”.* The government is continuing to work through those complexities, and as such I am looking forward to the recommendations of the Committee on the merits of introducing further mechanisms, drawn from the evidence base that you have gathered.

Yours sincerely,

Huw Irranca-Davies AS/MS
Y Dirprwy Brif Weinidog ac Ysgrifennydd y Cabinet dros Newid Hinsawdd a Materion Gwledig
Deputy First Minister and Cabinet Secretary for Climate Change and Rural Affairs