Applying behavioural insights to child protection: venturing beyond the low-hanging fruit

ANNALESE BOLTON*

School of Psychology, University of New South Wales (UNSW), Sydney, Australia and Neuroscience Research Australia (NeuRA), Sydney, Australia

BEN R. NEWELL

School of Psychology, UNSW, Sydney, Australia

SIMON GANDEVIA

Neuroscience Research Australia (NeuRA), Sydney, Australia and School of Medicine, UNSW, Sydney, Australia

JAMES PEEK

Family and Community Services, Behavioural Insights Unit, NSW, Sydney, Australia

ELENA BERROCAL CAPDEVILA

Family and Community Services, Behavioural Insights Unit, NSW, Sydney, Australia

Abstract: We explore whether simple behavioural insights techniques can be successful for addressing a policy issue within one of society’s more complex and difficult sectors: child protection. Child protection reporting practices in New South Wales, Australia, reveal that the public’s primary response is to report to the statutory authority, who only deal with cases of the highest risk. As a result, a large volume of statutory resources are spent processing lower-risk reports that lead to no benefits for lower-risk families and slow down response times for families that require a statutory response most. Our goal was to reduce lower-risk reporting by encouraging alternative responses to these situations. To do this, we altered report feedback for cases deemed lower risk in order to make alternative responses more salient and we added a persuasive message framed as a gain or loss. We then examined subsequent reporting accuracy. We found that our trial was linked to a modest improvement in reporting accuracy, though the results may have been diluted by a spill-over effect. We discuss how facilitating a greater behavioural change likely requires multi-organization collaborations, extending the range

* Correspondence to: Annalese Bolton, UNSW/Neuroscience Research Australia, School of Psychology, Matthews Building UNSW, Randwick, NSW 2052, Australia.
Email: a.bolton@unsw.edu.au
of insights drawn from behavioural science and/or addressing issues from multiple angles.

Submitted 3 July 2018; revised 12 April 2019; accepted 16 April 2019

Introduction

Over the last ten years, there has been an explosion of interest in applying simple techniques derived from behavioural science to public policy (e.g., OECD, 2017). These ‘behavioural insights’ have resulted in improved policy outcomes across countries and policy sectors (e.g., consumer protection, education, energy, environment, finance, taxes, telecommunications, etc.; OECD, 2017). However, such interventions have primarily targeted situations involving one-off behaviours with a binary desirable decision (i.e., the ‘low-hanging fruit’), such as with tax compliance (Sanders et al., 2018). Here, we ask whether these simple techniques can be successful for one of society’s more complex and difficult policy sectors: child protection.

Approaching such a complex, emotive policy area as this requires careful thought about the tractable areas for improvement and the potential tools available to the behavioural scientist. This paper documents that process in a study in which we conducted a randomized controlled trial (RCT) with the statutory child protection agency of the Government of New South Wales (NSW), Australia, with the goal of improving the public’s response to child abuse/neglect situations.

The public policy issue in a nutshell

The core aim of this study was to improve outcomes for vulnerable children and families by increasing the allocation efficiency of limited resources. Specifically, in 2015, the NSW FACS Behavioural Insights Unit (FACS BIU) identified the need for significant improvements in the ways in which people whose work brings them in contact with children (e.g., police, health professionals, education staff – so-called ‘mandatory reporters’) could report their concerns about cases of potential child abuse to the relevant authorities. Mandatory reporters are required in the NSW child protection system to monitor those children and respond appropriately. When a case is deemed sufficiently serious – defined as Risk of Significant Harm or ROSH – reporters are required to engage statutory child protection via making a report to the NSW Child Protection Helpline. The key issue identified by FACS BIU was that people used the Helpline as a primary response for all child protection situations rather than considering the various responses that might be more
Applying behavioural insights to child protection

Applying behavioural insights

Preliminary investigations by FACS BIU identified a range of factors that appeared to be contributing to this ineffective use of the Helpline. These factors included that mandatory reporters followed their ‘gut feeling’ rather than using the publicly available Mandatory Reporting Guide (MRG; which is an online interactive decision tree developed by the NCCD (2016)\(^1\) to assist people with identifying potential ROSH cases) and were unaware of alternative responses for the vulnerable families who do not meet the ROSH threshold. In addition, they also felt that the easy access to the Helpline made it the optimal response given competing commitments.

From a behavioural science perspective, the lessons from this preliminary investigation were relatively straightforward: mandatory reporters needed to be given usable information about alternative methods for responding, thus reducing the tendency to rely on the Helpline for all situations regardless of the level of risk involved. In addition, and perhaps as a means to motivate correct use of these alternative methods, reporters needed to be made aware of the potential impact – in terms of the general resource allocation issue – of simply relying on a ‘gut feeling’ rather than making a more considered and hopefully accurate response.

The primary conduit identified for conveying this new information was the feedback letter provided to mandatory reporters after making a report to the Helpline; specifically, the letters sent after a report was deemed not significantly serious and thus no further action was taken – defined as a Closed report. There have been many successful behavioural insights interventions involving feedback letters, or information letters in general (OECD, 2017), and although it was unlikely to be a panacea in this case, it provided a simple gateway for communicating more effectively with mandatory reporters.

An example of the feedback letter used by the Helpline before our trial began is shown in Figure 1(a). As can be seen, it does not provide mandatory reporters

\(^1\) The MRG was developed by the United States National Council on Crime and Delinquency, Children’s Research Center, in 2016.
Dear 

Thank you for contacting Community Services on 3 August 2016 in your capacity as a mandatory reporter about the following child/ren/ young person/s: 

Child/Children: DOB: 29/12/2011

The information you provided has been assessed and a decision has been made that it does not meet the statutory reporting threshold of 'risk of significant harm'.

You may have used the Mandatory Reporter Guide (MRG) prior to making your report to the Child Protection Helpline. The MRG, available at [www.KEEPTHEMSAFE.NSW.GOV.AU](http://www.KEEPTHEMSAFE.NSW.GOV.AU) aims to assist you in identifying the suspected level of risk to a child or young person.

The Child Protection Helpline applied a more comprehensive process to assess the information and it was determined that a statutory response is not warranted.

If you work within an agency that has a Child Wellbeing Unit, please contact this Unit to record your concerns and to discuss possible referral options.

Yours sincerely

Duty Team Leader
Child Protection Helpline.

---

**Figure 1.** (a) *Control* letter, (b) *New:Gain* letter.

with the information that the FACS BIU investigation suggested they wanted (e.g., why a particular decision had been reached, alternative pathways for responding), and there is no indication of the potential impact on resources of inaccurate reporting. The new letters developed for our trial aimed to overcome these issues by using behaviourally informed principles.
Dear [Name],

We acknowledge and thank you for the time and effort you have invested reporting your concern. However, in this instance:

**YOUR REPORT WAS ASSESSED AS NON-ROSH AND CLOSED WITHOUT ACTION**

By reducing the number of non-ROSH closed reports, up to 164 additional casework hours every day could be spent allocating resources to children and young people who would benefit from a FACS response the most.

You can maximise the chance of making ROSH reports and prevent further escalation of harm to children by reading the ‘Common reasons why reports do not meet the ROSH threshold’ and following the Response Guide provided. You are mandated by law to make a report to the Child Protection Helpline when you have reasonable grounds to suspect that the child/young person is at ‘Risk of Significant Harm’ (ROSH).

How did we come to our decision regarding your report?

To assess the risk, we combined the information you told us with information we hold about the child/young person and then applied a decision tool. The ROSH threshold was not reached and a statutory response is deemed not necessary at this time. Your report has been closed with no further action. This means that there will be no further assessment or action from FACS.

**Common reasons why reports do not meet the ROSH threshold:**

- When the information you provide us does not contain enough detail for us to act upon. For example, we need to know:
  - the current location of the child or parent
  - how the child was actually impacted by the incident
  - an explanation of what you mean by terms such as “sexual abuse” etc. using observable specific behaviours
- If the parents/caregivers are not causing or significantly contributing to the harm and are appropriately engaged in trying to address the issues of concern
- The risk was not high enough to meet the ROSH threshold which is a requirement for FACS involvement. Please note that the decision tool we applied to assess your report has a higher risk threshold than the Mandatory Reporting Guide.

**What can you do now to protect this child and best respond to children at risk in the future?**

You can help children get the most appropriate services at the earliest possible time by linking them with appropriate supports before their situation escalates to ROSH. Use the Response Guide provided to continue to protect this child/young person now and help you decide how best to respond to children at risk in the future.

Yours sincerely

Duty Team Leader (Child Protection Helpline)

---

Disclaimer: The information contained in this facsimile is privileged and confidential, and is intended only for the use of the addressee. The Child Protection Helpline has used its best endeavours to ensure that the information is correct. If you are not the intended recipient, you are notified that any disclosure, reproduction, distribution or other use of this communication is strictly prohibited. Confidentiality and legal privilege attached to this communication are not lost by reason of mistaken delivery to you. If you have received this communication in error, please notify the sender by reply transmission and destroy the facsimile without copying or disclosing it.

---

Figure 1b.
The three bottlenecks or impediments to accurate reporting – poor knowledge of alternatives, an unhelpful tendency to rely on the Helpline and framing of the impact – all lend themselves to interventions and techniques that are well established in the lab. The general debiasing technique of considering alternative responses (e.g., Arkes, 1991; Larrick, 2004; Soll et al., 2016) has shown great promise in leading people away from narrow thinking and improving judgement and decision-making (e.g., Anderson & Sechler, 1986;
Anderson, 1982; Arkes et al., 1988; Baron et al., 1988; Hirt & Markman, 1995; O’Brien, 2009). Simply providing people at the time of decision-making with different stimuli (e.g., alternative responses) that may not otherwise have been accessed can overcome the tendency to over-rely on associative memory for easily accessible responses (e.g., Arkes, 1991). To that end, the new letters presented information about the potential basis for the current outcome (i.e., a Closed report) and an accompanying Response Guide detailing the steps reporters should take when determining their actions next time they encountered a potential child abuse situation.

Providing alternative responses can act to reduce reliance on a particular response, but overcoming the status quo also requires people to be persuaded that a different response is necessary. Decision-makers often engage in strategies to minimize the cognitive/mental effort of thinking (e.g., Shah & Oppenheimer, 2008) and to avoid deciding by postponing, failing to act or accepting the status quo (e.g., for a review, see Anderson, 2003). This may be an issue if mandatory reporters perceive that the alternative responses are outside of their role due to either their own expectations or judgements of their ability, the training they have received, their current role description or even the expectations of their employers. To address these possibilities, we built on basic research into persuasion (e.g., the Elaboration Likelihood Model (ELM) of Persuasion; Petty & Cacioppo, 1986; Petty et al., 2004) in an attempt to formulate strong, persuasive arguments that would resonate with reporters. In the preliminary investigations by FACS BIU, reporters had commented that their reporting behaviour is motivated by what they consider to be in the best interests of children and young people. According to the ELM, when people are motivated, personally invested and have high ability, persuasion results from individuals carefully considering the information presented. In these situations, the strength of persuasive arguments is particularly important (Petty et al., 2004). Thus, to maximize engagement, the new letter attempted to increase mandatory reporters’ capacity to process the information by keeping the language as straightforward as possible, avoiding any jargon and providing all of the information required to engage in the alternative responses.

The final behavioural principle that we built on was the often-noted differences in how people react to outcomes framed as losses versus gains (e.g., Kahneman & Tversky, 1979). Persuasive arguments framed as losses (defined as bad things occurring and/or good things not occurring) have been shown to be more effective at promoting a desired goal than those framed as gains (e.g., Levin et al., 1998; Piñon & Gambara, 2005). Converging lines of evidence from research on goal framing indicates that a desired behavioural change itself can be considered risky or safe depending on the extent to which people perceive that a desired behaviour affords an unpleasant outcome (Rothman et al., 2006). When
people consider a behaviour where there is a risk of an unpleasant outcome, a loss-frame appeal should be more persuasive (as people are more risk-seeking when presented with loss-framed information). However, when people consider engaging in a behaviour where there is little risk of an unpleasant outcome, a gain-frame appeal should be more persuasive (as people are more risk-averse when presented with gain-framed information). In the current context, not making a report to the Helpline and instead relying on the alternative responses for lower-risk situations could be construed as a risky option. This is because failure to report to the Helpline when required can result in severe penalties (e.g., the removal of a reporter’s professional licence). As such, according to Rothman et al.’s (2006) conceptualization (see also the meta-analyses by O’Keefe & Jenson, 2007, 2009), a loss frame should be more effective at producing the desired behavioural response than a gain frame, as our mandatory reporters may be more willing to take this risk (i.e., risk-seeking behaviour in the face of a potential loss).

Although reporters might construe their reporting requirements as involving risks to themselves, our main aim in the messaging was to highlight the potential benefit to others of reporting accurately. Existing research suggests that in these other-focused contexts loss framing can lead to desired behaviour change (e.g., Fisher et al., 2008; Chou & Murnighan, 2013; Erlandsson et al., 2018). For example, Chou and Murnighan (2013) found that significantly more undergraduates donated blood when an appeal was presented with a loss frame (you can prevent a death) than a gain frame (you can save a life). In a second study, Chou and Murnighan found that presenting undergraduates with a charity’s appeal framed as either helping others avoid a loss (e.g., “You can minimize the possibility of a decrease in the charity’s ability to help John and others get out of unemployment by…”) or obtain a gain (“You can maximize the possibility of an increase in the charity’s ability to help John and others get out of unemployment by…”) resulted in significantly more helping behaviour. We thus predicted that loss framing would have a positive impact on the accuracy of reporting.

The new feedback letters

Putting all of these insights together resulted in the persuasive messages outlined in Table 1 and two new feedback letters – an example of the New: Gain letter is shown in Figure 1(b). The messages cut straight to the heart of the resource allocation issue by highlighting the additional number of caseworker hours that could be allocated towards processing reports (Gain) or the number of caseworker hours lost in processing sub-ROSH threshold reports (Loss). The other information in the letters was identical across the
Table 1. The persuasive message contained within the two versions of the New feedback letter.

<table>
<thead>
<tr>
<th>Feedback letter version</th>
<th>Persuasive message</th>
</tr>
</thead>
<tbody>
<tr>
<td>New:Gain</td>
<td>By reducing the number of non-ROSH closed reports, up to 164 additional casework hours every day could be spent allocating resources to children and young people who would benefit from a FACS response the most. You can maximize the chance of making ROSH reports and prevent further escalation of harm to children by…</td>
</tr>
<tr>
<td>New:Loss</td>
<td>Up to 164 casework hours are taken up every day assessing non-ROSH closed reports. This is time that we cannot spend allocating resources to children and young people who would benefit from a FACS response the most. You can minimize the chance of making non-ROSH reports and prevent further escalation of harm to children by…</td>
</tr>
</tbody>
</table>

Gain and Loss versions and the content was developed by FACS BIU in consultation with representatives from the Helpline, NSW Department of Premier and Cabinet BIU and academics. The New letter incorporated the additional information requested by mandatory reporters,\(^2\) as well as a three-step ‘Response Guide’. Modifications to the letters were made following consultation with two focus groups: one consisting of mandatory reporters (\(n = 5\)) and the other of child protection professionals from various organizations (\(n = 6\)).

**Trial design, baseline data and statistical power**

To examine whether the New letters (New:Gain and New:Loss conditions) were more effective at preventing avoidable future Closed reports than the standard letter (i.e., Control condition), we examined mandatory reporters’ subsequent reporting behaviour within the three-month trial period. As depicted in Figure 2, we examined whether mandatory reporters made another report following receipt of one of the feedback letters. For mandatory reporters who did not make a subsequent report during the trial, we recorded their report outcome as ‘No Call’. When reports are made, the Helpline makes a judgement as to whether a report meets ROSH, is Closed or is non-ROSH but

\(^2\) It was not possible to provide individualized feedback as the Helpline staff who generate the letters do not have access to the reasons individual reports do not meet the ROSH threshold and it was not possible to alter the system for this trial. As such, the new letters included common reasons as to why reports are classified as non-ROSH.
the information is forwarded\(^3\) (referred to as ‘Other’). As such, in our study there were four possible subsequent report outcomes: No Call, ROSH, Closed and Other.

Due to practical constraints (i.e., IT systems and infrastructure changes), the trial could only run from 11 July to 9 October 2016. Unfortunately, the cut-off for the follow-up period to examine subsequent reporting behaviour was 11 October 2016 (two days after the trial ended), as a revised version of the MRG was released on this date and it was reasonable to assume that this might impact reporting behaviour. Due to established seasonal patterns of reporting behaviour, a baseline sample (Baseline condition) was examined from the same period in the year prior to the trial (i.e. 11 July to 9 October 2015).

\(^3\)Some non-ROSH reports require action. For example, when situations involve a criminal offence, the information can be forwarded to the police; when there is an open child protection plan for the reported child/young person, the information is forwarded onto those managing the child’s case; or when the same incident is reported by multiple reporters in which a report meets ROSH and subsequent reports are classified as non-ROSH, but the information is added to the ROSH report.
As displayed in Table 2 and Figure 3, during the Baseline period, 3724 mandatory reporters made an initial Closed report for which they received a feedback letter. Of these, 2327 (62.5%) did not make a subsequent report during the trial period (i.e., No Call). Of those who did make a subsequent report, 645

<table>
<thead>
<tr>
<th>Subsequent report outcome</th>
<th>Baseline</th>
<th>Control</th>
<th>New:Loss</th>
<th>New:Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROSSH</td>
<td>645</td>
<td>209</td>
<td>189</td>
<td>178</td>
</tr>
<tr>
<td>Other</td>
<td>265</td>
<td>69</td>
<td>51</td>
<td>70</td>
</tr>
<tr>
<td>Closed</td>
<td>487</td>
<td>135</td>
<td>105</td>
<td>111</td>
</tr>
<tr>
<td>No Call</td>
<td>2327</td>
<td>706</td>
<td>749</td>
<td>722</td>
</tr>
<tr>
<td>Total</td>
<td>3724</td>
<td>1119</td>
<td>1094</td>
<td>1081</td>
</tr>
</tbody>
</table>

Table 2. Number of mandatory reporters who made a Closed report for which they received a feedback letter by letter condition and subsequent report outcome.

*a These represent final n as used in the data analysis.

As displayed in Table 2 and Figure 3, during the Baseline period, 3724 mandatory reporters made an initial Closed report for which they received a feedback letter. Of these, 2327 (62.5%) did not make a subsequent report during the trial period (i.e., No Call). Of those who did make a subsequent report, 645

Figure 3. Subsequent reporting behaviour after being sent a Closed report feedback letter.
(17.3%) made a ROSH, 487 (13.1%) made a Closed and 265 (7.1%) made an Other report.

From these baseline data, we predicted that the N for our trial would roughly equal 3724, with an approximately equal n for each condition (i.e., 1241 in each of the Control, New:Gain and New:Loss letter conditions). We predicted that the outcome proportions in the Control condition would roughly equal the Baseline proportions (i.e., No Call ≈ 0.6249/n ≈ 776; ROSH ≈ 0.1732/n ≈ 215; Closed ≈ 0.1308/n ≈ 162; Other ≈ 0.0712/n ≈ 88).

As far as we are aware, there have been no gain/loss-framing studies conducted within the domain of child protection response decision-making (or, for that matter, within the broader field of child protection). Moreover, existing literature on goal framing shows a high variability of effect sizes, even between studies of very similar situations (Piñon & Gambara, 2005; O'Keefe & Jenson, 2007, 2009). As such, it was difficult to predict the magnitude of the impact of our interventions in the current situation. To err on the side of conservatism, we thus attempted to have enough power to detect a small effect size of a 3% difference between the proportions of subsequent reports in the Control condition and the trial condition in which we predicted the effect to be smallest (i.e., New:Gain). Using a standardized effect size formula for a \( \chi^2 \) test, assuming Control showed similar proportions to Baseline with 2400 cases (i.e., an approximate \( n \) for the comparison of two conditions), indicated that we had greater than 90% power to detect a difference of this magnitude (see Supplementary Materials for details of this calculation).

Hypotheses

One hypothesis was that simply alerting mandatory reporters of the available alternative responses via a New letter (New:Gain and New:Loss) would increase their accuracy at determining when to contact the Helpline. According to this hypothesis, we would expect that, compared to those who received the standard letter, for those who received a New letter, the proportion of subsequent:

- Closed would be significantly smaller;
- No Call would be significantly larger; and
- ROSH would either remain the same or increase.

Additionally, we hypothesized that the persuasive message framed as a loss would result in significantly greater accuracy in future reporting behaviour for those who received a New:Loss letter than a New:Gain letter (i.e., the differences between standard letter condition and New:Loss would be greater than the differences between standard letter condition and New:Gain). It
was unknown what impact the letters would have on Other due to the eclectic mix of issues that fall into this category.

Although we did not have specific hypotheses about whether different profession types (e.g., education, police, health, non-governmental organizations and other) might respond differently to the New letters, we were able to explore whether this functioned as a covariate.4

A potential – and undesirable – alternative hypothesis was that there would be a general reduction in reporting behaviour irrespective of the level of perceived risk. If this hypothesis were supported, we would see a significantly smaller proportion of subsequent reports in all report categories (ROSH, Closed and Other) for those who received a New compared to the standard letter, and the number of ROSH would be less than expected from the historical data. Stopping mandatory reporters from contacting the Helpline when they thought children were at risk of significant harm would be problematic due to the serious risks to vulnerable children. Therefore, we took precautions to protect against this outcome. These included attaching highly visible reminders to all letters sent during the trial that reporters were mandated by law to report suspected ROSH (see Supplementary Material for the email text accompanying the trial letters). During the trial, we also monitored the total number and proportion of ROSH (for all reporters, not just the subset involved in the trial, as these data were more readily accessible) and compared these data with forecasts. It was predetermined that if the number and proportion of ROSH fell below the forecast, the trial would be ceased.

Due to these precautions, the representativeness of our Control (standard letter) condition as ‘behaviour as usual’ may have been undermined. In addition, our participants were aware that they were part of a trial and therefore may have modified their behaviour in response to their awareness of being observed (e.g., McCambridge et al., 2014). To address this issue, we included the Baseline as a quasi-control, as this group’s behaviour represents behaviour as usual without the interference of knowing they were part of a trial. Caution is recommended in interpreting comparisons with the Baseline as it was not included as part of the randomization procedure and we cannot rule out the possibility that a difference between our trial conditions and Baseline may be due to history effects or data variability over time. To mitigate against these possibilities and to evaluate the usefulness of the Baseline in assessing the effectiveness of the trial, we compared the composition of the Baseline and

4 The only reliable and meaningful demographic information we had access to regarding our participants was their profession type (most data routinely collected by FACS concern the vulnerable child and their family, rather than the reporter).
Trial participants in terms of the relative distribution of profession types. The results of these tests are reported in the Supplementary Materials and reveal that the Baseline was comparable in the proportions of professions. These analyses thus suggest that although it was not part of the randomized trial, comparisons to the Baseline are of some value in interpreting our results.

Method

Ethics approval was obtained from the University of Wollongong HREAC (approval HE16/234) for the trial and UNSW School of Psychology SCI HREAPC (approval 2751) for the data analysis.

Design

A RCT design was used whereby mandatory reporters who made an initial Closed report to the Helpline and received a feedback letter between 11 July and 9 October 2016 (the trial period) were randomly allocated to receive one of the three letters (Control (i.e., the standard letter), New:Gain or New: Loss). The data required to monitor mandatory reporters’ subsequent reporting behaviour are routinely collected. For this trial, mandatory reporters’ subsequent behaviour was defined as the outcome of their next report to the Helpline within the trial period after being sent a trial feedback letter.

Randomization

Due to practical constraints, randomization occurred at the report level, not by mandatory reporter (i.e., there was no practical way to flag mandatory reporters who had already received a trial letter). Therefore, each time a mandatory reporter made a Closed report during the trial period, they were randomly assigned to receive one of the trial letters. However, there was a lag time between the date the report was made and the date the feedback letter was sent (e.g., between a day to several weeks). As such, based on their reporting behaviour, some mandatory reporters received multiple versions of the trial letters before they made – or could make – a subsequent report (data for these mandatory reporters were excluded, as outlined below). As detailed in the Supplementary Materials, several measures were taken to ensure that the randomization process was implemented correctly.

Data collection, merging and inclusion and exclusion criteria

Data relating to the letters (i.e., letter condition, date letter was sent and report identification number) were manually collected from the Helpline by FACS BIU. These data were merged (using the report identification number) with
report outcome data routinely collected by FACS (all reports made to the Helpline by date of report, report outcome, report identification number, mandatory reporter’s identification number and reported child/young person’s identification number). As a single report can involve multiple children/young people, data for only one child/young person per report were included (e.g., children within the same household are generally allocated the same risk level: that of the highest risk identified within the household).

During the merging process, data for 240 feedback letters (i.e., 4.3% of all letters sent) could not be matched with a mandatory reporter (likely due to errors in recording the identification numbers). It is unknown which, or how many, mandatory reporters received these letters. As displayed in Figure 4, data for 3445 mandatory reporters met the inclusion criteria for this trial (i.e., the data indicate they were sent a trial letter for a Closed report). Of these, we excluded data for 151 mandatory reporters as they received multiple trial letter versions before they made – or could make – a subsequent report. As indicated in Table 2, the final n for the trial was 3294. Therefore, the total N (trial plus Baseline) was 7018.

Data analysis

Forecasts for the total numbers and proportions of ROSH and Non-ROSH (i.e., Closed and Other combined) were derived in SPSS from linear regression models using a moving centre average technique based on the same seasonal period data for the previous five years (i.e., 11 July to 7 August and 8 August to 4 September from 2010 to 2015).

Data were analysed in R using a generalized additive model (GAM; Hastie & Tibshirani, 1990) assuming a multinomial distribution and controlling for multiple comparisons using the Holm–Bonferroni method (code provided in the Supplementary Materials).

Results

Analysis of subsequent report outcomes

Figure 3 depicts the proportions of mandatory reporters’ subsequent report outcomes for those in the trial (Control, New:Loss and New:Gain) and Baseline conditions.5

5 During the trial, there was a minor wording change to the Response Guide provided with the New feedback letters. Statistical analysis indicates that this did not significantly impact the outcomes (GAM: F = 7.4679, p = 0.2797).
To examine the impact of the New letters, we compared the subsequent reporting outcome proportions. Statistical analysis indicates that there were significant differences between letter conditions (GAM: \( F = 28.27, p < 0.0027 \)). Profession type was not a significant covariate (GAM: \( F = 35.57, p = 0.0604 \)).

Figure 4. Data flow diagram.

To examine the impact of the New letters, we compared the subsequent reporting outcome proportions. Statistical analysis indicates that there were significant differences between letter conditions (GAM: \( F = 28.27, p < 0.0027 \)). Profession type was not a significant covariate (GAM: \( F = 35.57, p = 0.0604 \)).
After controlling for multiple testing using Holm-Bonferroni, compared to Control, the New:Loss letter resulted in a significantly larger proportion of subsequent No Call responses (p < 0.0313). Crucially, there were no significant differences in the proportion of ROSH between trial conditions (i.e., p > 0.05). All other comparisons with the Control condition were non-significant (i.e., p > 0.05). Such data provide some support that the New:Loss letter increased reporting accuracy.

Due to our concerns that the Control condition may not have represented ‘behaviour as usual’, we also examined differences between our trial conditions and Baseline (our ‘quasi-control’). There were no significant differences in the subsequent reporting outcomes between the Baseline and Control (i.e., p > 0.05), lending further support to the notion that these two conditions were reasonably equivalent. Compared to the Baseline, the New:Loss letter resulted in a larger proportion of subsequent No Call responses (p < 0.0013), as well as a significantly smaller proportion of Closed (p < 0.0065) and Other (p < 0.0085) responses. In addition, compared to the Baseline, the New:Gain letter resulted in a significantly larger proportion of subsequent No Call responses (p < 0.0390) and a significantly smaller proportion of Closed responses (p < 0.0421). Once again, crucially, there were no significant differences in the proportion of ROSH between Baseline and trial conditions (i.e., p > 0.05). These data lend somewhat stronger support to the notion that both New letters — though particularly the New:Loss letter — increased reporting accuracy to perhaps a greater degree than we were able to capture in our trial.

The impact of the trial on reporting behaviour

The magnitude of difference in subsequent non-ROSH reporting (i.e., Closed and Other reports combined) between the New:Loss and the Control conditions for participants in our trial was 3.94%; the difference between New:Loss and Baseline was slightly larger at ~5.94% (see Figure 3).

It is difficult to translate this difference to the impact the new letter had on overall reporting, as the frequency with which mandatory reporters are exposed to child abuse/neglect situations varies significantly between mandatory reporters and across time (e.g., some mandatory reporters may only be exposed once in their career, whereas others may have frequent exposure; changing the behaviour of those with frequent exposure may have a cumulative impact on overall reporting). Due to practical restrictions, we were unable to explore the impact of the trial letters on the cumulative behaviour of frequent reporters involved in the trial. However, we were able to undertake a broader examination of overall reporting behaviour across time. For example, during the trial, we monitored the overall number and proportion of ROSH. As
indicated in Table 3, the number of ROSH was within the forecast, but the proportion of ROSH was significantly higher than forecast during both time intervals, by between 2% and 3%.

Given such a short time interval, it is difficult to determine whether this increase was a direct effect of the trial or whether overall reporting behaviour is so variable that the actual proportion of ROSH regularly falls outside of the forecast confidence interval. To address this issue, we plotted the pattern of report outcome proportions over a longer period using a statistical quality control approach (Western Electric Company, 1956). Rather than examining ROSH versus Non-ROSH for all of the reports as we had in Table 3, we examined only reports made by mandatory reporters, separated Non-ROSH into Closed and Other and examined the data on a weekly basis (rather than monthly). We intended to plot data for the last five years (from when the ROSH/Non-ROSH distinction was introduced); however, from July 2014 onwards, there was a change in the way Other was categorized. Therefore, as displayed in Figure 5, we used weekly data from 1 July 2014 until the start of the trial on 11 July 2016 to calculate a centre line (i.e., average proportion) and upper and lower limits (i.e., three standard deviations above and below the average, respectively). The centre line and limits were then applied

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of ROSH reports</th>
<th>Proportion of ROSH reports</th>
<th>Year</th>
<th>No. of ROSH reports</th>
<th>Proportion of ROSH reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>6890</td>
<td>49%</td>
<td>2010</td>
<td>8551</td>
<td>51%</td>
</tr>
<tr>
<td>2011</td>
<td>7275</td>
<td>45%</td>
<td>2011</td>
<td>8377</td>
<td>46%</td>
</tr>
<tr>
<td>2012</td>
<td>7111</td>
<td>36%</td>
<td>2012</td>
<td>8370</td>
<td>37%</td>
</tr>
<tr>
<td>2013</td>
<td>9021</td>
<td>48.2%</td>
<td>2013</td>
<td>9981</td>
<td>46%</td>
</tr>
<tr>
<td>2014</td>
<td>9677</td>
<td>47.9%</td>
<td>2014</td>
<td>10,687</td>
<td>49%</td>
</tr>
<tr>
<td>2015</td>
<td>10,290</td>
<td>48.3%</td>
<td>2015</td>
<td>10,934</td>
<td>49%</td>
</tr>
<tr>
<td>2016: predicted</td>
<td>9980–11,902</td>
<td>46.1–48.9%</td>
<td>2016: predicted</td>
<td>10,214–12,958</td>
<td>44.8–51.0%</td>
</tr>
<tr>
<td>2016: actual</td>
<td>10,169</td>
<td>52.0%</td>
<td></td>
<td>12,497</td>
<td>53.3%</td>
</tr>
</tbody>
</table>

Table 3. Total number$^a$ and proportion of ROSH reports during the trial period.

$^a$ Data are for all reports, not just those made by mandatory reporters.
$^b$ The proportion of total reports that were ROSH.
$^c$ 95% confidence interval using moving centre average technique.

6 In July 2014, frequent absconder reports became its own category rather than being contained within the category of Other.
Figure 5. Weekly report outcome proportions for (a) ROSH, (b) Closed and (c) Other. Data are for reports made by mandatory reporters only. The centre line (CL) represents the average proportion of the pre-trial calibration data, while the upper and lower limits (UCL and LCL, respectively) represent three standard deviations above and below the CL, respectively. The UCL and LCL are variable. Black dots are observations that are more than three standard deviations above or below the CL. Grey dots are observations that violate one of the other Shewhart rules (i.e., two out of three consecutive points are on the same side of the center line; four out of five consecutive points are beyond one standard deviation on the same side of the center line; eight consecutive points on the same side of the center line). Note: UCL and LCL are variable.

*Data are for reports made by mandatory reporters only, weekly from 1 July to 31 December 2016. Black dots are observations that are three standard deviations above or below the center line. Grey dots are observations that violate one of the other Shewhart rules (i.e., Two out of three consecutive points are on the same side of the center line and are two standard deviations from the center line; Four out of five consecutive points are beyond one standard deviation on the same side of the center line; Eight consecutive points on the same side of the center line). Note: UCL and LCL are variable.
to weekly data during and after the trial period. Points that fall above the upper or below the lower limit are considered significant deviations and are represented by a black dot.

Figure 5 reveals that the proportion of report outcomes does vary across time, although it is contained within defined margins. For example, between 1 July 2014 and 11 July 2016, the weekly proportion of Closed is consistently between 23% and 30%, and the proportion of ROSH is between 48% and 56%. When data are grouped in monthly intervals, the pre-trial margins reduce to 25–28% and 50–54%, respectively.

The pattern of weekly report outcome proportions as displayed in Figure 5 demonstrates a noticeable change around the time of the trial. In particular, the frequency of black dots (which indicate that the weekly proportion differs significantly from the pre-trial weekly average) increases from during the trial to the post-trial period. Consistent with increased accuracy of reporting, the weekly proportion of ROSH trends upwards, Closed trends downwards, while Other remains relatively stable. Just prior to the trial, ROSH is significantly different from the average and Closed is close to significance. It is unclear why this occurred just prior to the trial (it could be due to random variation, although it may also have occurred from inadvertent influences of professionals who became aware of the trial goals in May 2016 and who function as intermediaries between mandatory reporters and the Helpline).

The overall pattern is consistent with an increase in reporting accuracy linked to the trial and may reflect an ongoing impact of the feedback letters after the trial period. Using the data from Figure 5, when comparing the average weekly proportion between the pre-trial and trial period, there is an overall reduction in the proportion of Closed by approximately 3%. Although this may seem small at first, it takes approximately 1.0–1.5 casework hours to process a single report. In 2016, there were 38,148 Closed reports made by mandatory reporters. With these figures, reducing the number of Closed reports even by 3% is equivalent to 1144–1716 casework hours per year. As the number of reports made to the Helpline increases each year (e.g., Australian Institute of Health and Welfare, 2017), the savings would also increase.

Discussion

We set out to test whether simple behavioural insights techniques, such as assisting people to consider alternatives and altering the valence of a persuasive message, could be effective at increasing desired behaviour when applied to one of society’s most complex and difficult policy sectors: child protection. Promisingly, we found that these techniques did lead to a statistically
significant real-life behavioural change in the desired direction. Specifically, we found evidence in support of the hypothesis that the New letter, particularly the New:Loss version, increased reporter accuracy. There was no evidence to suggest that the New letters induced mandatory reporters simply not to call the Helpline regardless of risk. However, the evidence for an increase in reporting accuracy was stronger when comparisons were made with our Baseline condition than our Control. This is, of course, potentially problematic because the Baseline was not part of our RCT and we cannot rule out the possibility that a difference between our trial conditions and Baseline may be due to history effects or data variability over time. That said, the balance checks for professions and the lack of any significant difference between our Control and Baseline conditions suggest that these two conditions were reasonably comparable.

One reason why the comparisons between the trial conditions and Baseline were larger than with Control is that a ‘spill-over’ effect of the trial may have occurred. Due to the complexity of issues that mandatory reporters consider when making a report, it is reasonable to assume that our participants consulted their colleagues and/or management about their decision to report. For example, Weinstein et al. (2000) found that, before making a child protection report, 90.3% of health professionals (even highly experienced professionals) consulted a supervisor or colleague. In our study, a spill-over effect may have unintentionally occurred if mandatory reporters consulted with someone previously exposed to the new trial letters. This is particularly likely in organizations that are frequent reporters or where staff routinely contact professionals who function as intermediaries between mandatory reporters and the Helpline. This kind of spill-over effect would have introduced contamination between trial conditions (e.g., a mandatory reporter in the Control condition may have been influenced by someone exposed to another condition), thereby possibly diluting differences between the three trial conditions. Indeed, during the trial, a mandatory reporter who had only received the standard Control letter contacted the Helpline requesting a copy of the Response Guide (which had only been included with the New letter). This suggests that those in the Control condition may have actively sought information provided in the other conditions, further supporting suggestions of a spill-over. Unfortunately, FACS BIU closed in December 2017 (prior to the finalization of the data), and therefore no follow-up or further investigation was possible. However, future studies may benefit from including a post-trial survey in which participants are asked about the processes they engage in when they respond to situations, which could elucidate whether spill-over was potentially occurring.
These issues notwithstanding, our broader analysis suggested that the trial was linked to an increase in reporting accuracy across time. Pinpointing the exact reasons for this increase is difficult, as it often is in fieldwork where compromises have to be made (Bolton & Newell, 2017). For example, completely beyond the control of this research team, a revised version of the MRG decision tree tool was released on the day after the trial ended. The revised MRG incorporated the steps and links to alternative services outlined in the Response Guide that we developed to accompany our new trial feedback in order to highlight the available alternative responses (see Figure 1(b)). As such, it is impossible to tell whether the pattern of increased reporting accuracy post-trial was due to the ongoing impact of our trial letters, the MRG revisions or simply that the Response Guide is now accessible to all mandatory reporters. While this inability to pinpoint cause is frustrating from the behavioural science perspective, the improvement in reporting and hence the positive impact on the policy concern is encouraging.

Overall, our results provide evidence that assisting people to consider alternatives (e.g., Larrick, 2004) and altering the valence of a message frame (e.g., Levin et al., 1998) can produce detectable real-life desired behavioural change, even for complex social policy issues. These results thus provide an encouraging extension of predominantly lab-based findings to the field, and to a new policy area. Our study is a modest first step, however. Facilitating a greater behavioural change in mandatory reporters from reporting to responding may require a more intensive approach than changing the feedback letter. For one, it may be that studies addressing complex social policy require multi-organization collaborations. This approach allows greater buy-in from all stakeholders involved and permits a more holistic understanding of what is in society’s best interest.

In conjunction with this multi-organization approach, rather than trying to facilitate a complex behavioural change by providing feedback after mandatory reporters have responded, an alternative method might be to facilitate mandatory reporters’ (and the community’s) understanding of appropriate responses before they need to respond. This may include developing training programmes using insights from cognitive and educational psychology and conducting RCTs to examine which type of training is most effective at producing appropriate responding (at all risk response levels, not just whether the risk meets the statutory significance threshold). Indeed, FACS BIU developed a scenario-based training programme that they piloted with 180 mandatory reporters from non-governmental organizations, police and independent schools, and some of the current authors have begun collaborating with the FACS Helpline and NSW Health Child Wellbeing Unit to develop a training programme aimed at bolstering allied health professionals’ ability to identify appropriate responses.
Conclusion

This study demonstrates that the behavioural insights field can venture beyond the low-hanging fruit, which has been its primary playground over the last ten years. Like any ten-year-old, the potential is undeniable, but there is still a lot to learn. For instance, perhaps facilitating a greater behavioural change for complex social policy issues requires a more intensive approach, either via multi-organization collaborations, extending the range of insights drawn from the behavioural sciences and/or addressing issues from multiple angles.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1017/bpp.2019.12

Acknowledgements

This research has been initiated by, and undertaken in collaboration with, the NSW Department of Family and Community Services. However, the information and views contained in this study do not necessarily, or at all, reflect the views or information held by the NSW Government, the Minister for Family and Community Services or the Department.

We thank UNSW’s Stats Central consultants Dr Nancy Briggs and Dr Gordana Popovic for assisting with data analysis; FACS Research and Analysis staff Soemya Agarwal and Nik Nimpradit for providing feedback on the trial design and assisting with the forecasts; and NSW Department of Premier and Cabinet Behavioural Insights Unit for their initial advice on the trial randomization methodology.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. Annalese Bolton is funded via the Australian Government Research Training Program Scholarship.

References


Sanders, M., V. Snijders, and M. Hallsworth (2018), ‘Behavioural science and policy: Where are we now and where are we going? Behavioural Public Policy, 2, 144–167.


