# Asia Wildlife Enforcement and Demand Management Project: Activity 5.5

## Briefing Paper:

## Strengthening Demand Reduction Measurement: Options on methods from behavioural science

James Walsh and Sonja Vogt[1]

[1] The authors would like to thank James O'Brien, Elaine Marie, Cédric Grützner, and Lauren Harrison for their help in preparing this paper.

# Contents:

# Introduction

The illegal trade in wildlife threatens the survival of many species and has devastating and wide ranging economic, social, and environmental consequences. Poaching and trafficking also finances corruption, undermines the rule of law and good governance, and drains States and communities of their natural capital and cultural heritage. Historically, efforts to prevent this trade have focused on efforts to restrict supply, for example through policies, regulations, legislation, and the effective enforcement of these various 'societal controls'. More recently however, there has also been recognition of the need to complement this work through efforts to reduce consumer demand, using measures such as messaging to shape individual motivation, based on best practice in behavioural science. At the 17th Conference of Parties (J'burg, Sept 2016) to the Convention on International Trade in Endangered Species of Fauna and Flora (CITES), a Resolution on Demand Reduction was adopted, recognising the importance of such well targeted, species-specific and evidence-based campaigns, that aim to engage key consumer groups and target the motivations for the demand, to change consumer choice and buyer behaviour[2].

Commensurate with this recognition of the importance of demand reduction approaches, there has been increased scrutiny around the methodologies being applied to inform initiative design and assess progress and impact. The following excerpt from a recent CITES Secretariat report (SC69 Doc 15[3]) summarizes key issues identified as arising:

*Public opinion surveys are often used to measure the effectiveness of a demand reduction campaign. There are many factors that can affect the accuracy and usefulness of a public opinion or consumer response poll, such as the sample size, the audience you choose and the likelihood of the respondents telling the truth. Demand reduction and behaviour change are often a gradual process. Although it is important to measure the impact of a campaign, the Secretariat is of the opinion that Parties should be cautious when they make statements on the achievements of a demand reduction campaign, particularly when a claimed dramatic drop in demand is not echoed by a change in the level of poaching and smuggling, in which case such claims will risk diminishing the credibility of demand reduction strategies.*

Within this frame, Activity 5.5. of the *Asia Wildlife Enforcement and Demand Management* project, supported by the EU and implemented by TRAFFIC via CITES MIKE, has adopted a focus of gathering data around current and best practice, and of suggesting suitable methodologies to help improve demand reduction impact measurement. This focus has segued neatly with TRAFFIC's wider engagements in, and support for, the CITES process around demand reduction, pursuant to Decision 17.48 (see here and here). Thus, using support from CITES MIKE, with complementary contributions from USAID and GIZ, early activities in relation to this work have included research into what is working and what isn't in relation to demand reduction initiative implementation. Mechanisms have included through CITES Notification 2018/15, as well as associated interviews and desk-based research. The latter specifically

---

[3] https://cites.org/sites/default/files/eng/com/sc/69/E-SC69-15.pdf

sought to build on analogous material, such as the Defra 'Research Analysis' (here), SBCC Community of Practice engagements at the *1st International Conference on Behaviour Change to Tackle Illegal Wildlife Trade* (here), and the Expert Roundtable discussions and process generating the *SBCC M&E Good Practice Guidelines* here. TRAFFIC collated and synthesized key insights emerging in the Report here for the CITES Secretariat. A draft was presented and discussed at an 'Expert Workshop', convened by the CITES Secretariat, Chaired by Thailand, and with 11 participating Parties, in Bangkok (27th November, 2018).

To complement these CITES focused activities, high-level experts in behavioural science research methodologies were also engaged, to draw together 'options' for presentation, consideration and discussion, at a subsequent 'Deep Dive Working Group' during the *2nd International Conference on Behaviour Change to Tackle Illegal Wildlife Trade.* This paper is associated with this complementary process, and summarises the options and some associated methodological reflections, through a focus on the following key themes:

1) **Causal Inference:** Interventions are generally not designed and assessed in ways that allow conclusions that causally attribute demand reductions to interventions efforts. Understanding whether and why interventions work is essential to efforts by the community as a whole to build a robust evidence base. Markets are complex and multi-dimensional, making it challenging to understand whether efforts had any impact at all. Therefore, it is crucial that interventions and impact assessments are carefully designed.

2) **Impact Measurement:** As wildlife trade is illegal in most places, it is especially challenging to elicit reliable measures from people who demand these products. Simple self-reported surveys are often ineffective at distinguishing whether people do not demand a product, do not want to report illegal behaviour, or do not want to report a preference for illegal wildlife to a wildlife NGO. Innovative measurement efforts to overcome these challenges can and should be used to get a more accurate measure of demand. Otherwise, it will be difficult to know what data tells us.

Next steps following production of this paper, CITES report, 'Expert Workshop' and; 'Deep Dive Working Group' discussion, are the development of a methodology to help move current practice towards best practice, in demand reduction impact measurement. The methodology will be discussed with a panel of experts and practitioners, before being considered for potential inclusion in the following guidance and capacity building materials being prepared to support governments, NGOs and others, implement demand reduction initiatives:

- CITES Guidance, associated with Resolution 17.4;
- UNODC training resources / content for the World Wildlife Crime Report
- Knowledge Bank resources for GEF 6/7 Global Wildlife Program countries
- Materials for inclusion in the Wildlife Consumer Behaviour Change Toolkit (www.changewildlifeconsumers.org), including the M&E components of a MOOC

# Causal Inference

A key challenge identified in current practice in relation to demand reduction impact measurement, is the extent to which the intervention is the cause of any change —not just that it correlates with it. Drawing from other approaches in behavioural science, to establish whether an intervention causes behavioural change, we must imagine what happened if the intervention had never taken place (Angrist and Pischke 2008). Would a target group have purchased rhino horn if our organisation had not created and promoted the campaign we just implemented? Would the target group have disapproved of the use of ivory had the intervention not taken place? The imagined alternative state can be considered a "counterfactual." Some of the challenges associated with identifying attribution of impact / associated counterfactuals, are considered further in this section.

Tracking the same individual over time (i.e. before and after the intervention) is a very common approach in sectors beyond illegal wildlife trade. This method is generally more challenging within illegal wildlife trade however, due partly to the inherently sensitive nature of the survey subject matter, but also partly due to shifting social acceptability around IWT consumption behaviours. As a result, it may be very challenging to secure the same individual in the base- and endline study, and in addition, to tell whether individuals participating in the baseline survey changed their behaviour because of the intervention under study, or because of shifts in social pressure and acceptability, or some other factor. The potential for "contamination" of impact is additionally important in contexts where many NGOs are running similar interventions.

An alternative strategy to consider therefore, is to compare the behaviour of individuals exposed to the intervention to a control group. This approach is methodologically robust but may be challenging to implement. Unless extraordinary measures are taken to ensure comparability, intervention groups may differ systematically in ways that make them an inadequate counterfactual. For example, demand reduction interventions may take place in a particular region for a deliberate reason (e.g. because demand is higher in the area or the target group is less challenging to reach due to an urban, rather than rural, sample). Individuals may also have self-selected into the intervention (if they are interested in wildlife protection, for example). Such challenges are called selection bias. While it may be possible to account for some observable differences between the groups (i.e. differences for which you have data – for example, gender or region), there may be many unobservable differences that you cannot account for (e.g., they may have different beliefs about animals—or some other factor you never thought to measure).

In summary, the most common approaches to monitoring and evaluation often do not adequately identify whether interventions caused the change they were intended to produce.

## *Identifying causal effect*

Randomised Controlled Trials (RCT) offer a potential solution to this challenge and represent a standard method for testing the causal effect of an intervention used in the field of medicine and public health, for many decades. Since the 2000s, the method has also become a ubiquitous tool in development economics and policy (Takavarasha and Glennerster 2013).

When intervention designers randomly assign an individual (or group) to either the treatment or control status, the likelihood that a given person ends-up in one group or the other is essentially pure chance. It does not matter what region they are in, what their income is, whether they love or hate animals; they are just as likely to be in either group. The implication of this is that if enough people participate in the intervention (and serve as a control group), the two groups will be statistically comparable. For example, we expect that roughly the same number of men and women or those on high or low incomes, will be in each group. This technique eliminates the concerns outlined in the previous section that observed changes might not be attributable to the intervention. When the project is designed in this way, the intervention and control groups can be thought of as counterfactuals.

Once you have identified what interventions you want to test, you can decide on a way of measuring the desired behavioural change. After this, the process of an RCT typically follows these seven steps[4]:

**Step 1: Select a group to work with.** In some cases, you may want to take a random sample from the population. This allows you to estimate how the intervention would work if it was scaled up to the population as a whole. Often this will not be possible, for example, if you are targeting a population that is ambiguously defined (e.g., prospective consumers of an illegal wildlife product). It may then be necessary to take a convenience sample.[5] This is common practice in social science.

**Step 2: Collect basic data on your population.** You need to have a measure of the outcome you're interested in investigating. This should be collected at the baseline and then again at the end line. This must include participant identification and may include data that provides a descriptive account of the sample you are working with. Example descriptive data could be gender, age, education level, nationality, religion. This kind of data can help you ensure that the randomization produces groups that are balanced on important demographic variables and allows you to explore whether your intervention affected these groups differently (this is called heterogenous treatment effects).

---

[4] When implementing a randomised control trial, we recommend working with either a researcher or an organization with expertise in these methods.

[5] A "convenience sample" is a way of selecting study participants that emphasises ease of access over ensuring that they are representative of the population as a whole. For example, a study that uses data collected from people outside a grocery store might be a convenience sample.

**Step 3: Select the unit of randomization.** Generally speaking, there are two ways to randomise – at the individual level or the group level. The most appropriate level of randomisation will differ from intervention to intervention, depending on the research question and the design of the intervention in question. For some studies, it may be possible to randomise at the individual level (e.g., you put each person into a lottery, and assign them to a treatment and intervention group). For others, it may be necessary to randomise at the group level (e.g., you put groups of people – for example, classes, schools, villages – into a lottery and assign groups to the different statuses). We randomise at the group level when we are concerned that individuals in the treatment group will interact with individuals in the intervention group, and the effect of the treatment will "spill over" to the intervention group. If this happens, then we might underestimate how effective the treatment was.

**Step 4: Conduct a power analysis.** An evaluation has to be sensitive enough to statistically detect differences in the outcome measure between the treatment and intervention groups. To ensure that it is capable of doing this, you may want to assess the sample size necessary to detect real effects of the intervention. This aspect of the evaluation is called the "statistical power" of the study. From a logistical perspective, the primary way to improve study power is to increase the sample size.  The larger the sample size, the more powerful the study. Budgetary constraints typically prevent studies from getting too large.

**Step 5: Randomly assign each unit to either the treatment or control group.** The central part of an RCT is the randomisation, and it is a surprisingly straightforward process. You simply create two conditions (treatment and intervention) and then assign the unit of analysis to each condition at random. This procedure can be done in Excel or through statistical programming software such as Stata or R. There are different ways to randomly assign units to control and treatment groups. You can do it with a simple lottery design, i.e. a genuinely random assignment without any constraints. Often policy considerations call for the randomisation to be implemented in a particular way. For example, an intervention may be phased-in at random or rotated throughout a population at random, or people might be encouraged at random to participate in the treatment group but not in the control group.

**Step 6: Implement the intervention only for the unit in the treatment group.** It is unethical cannot force people to comply with our treatment. In cases where a significant proportion of the treatment group do not "comply" (e.g., do not receive the treatment), you can analyse the results from the perspective of Intention to Treat (ITT) – where you compare the whole treatment group against the whole control group, regardless of whether or not they actually a received the treatment. The assumption, in this case, is that ITT measures a lower bound of the intervention's impact.

**Step 7: Compare the average score on the variable of interest in the treatment and control groups.** This difference between the two scores provides the average treatment effect. In further analysis, you may want to investigate the effect of the intervention on different segments of the sample or look at whether the intervention's effect changed across time (which you would need more data for).

<div style="border:1px solid black; padding:10px;">

**Box 2. Considerations for choosing your unit of analysis**

**Type of intervention:** If your intervention is highly social in nature, the best randomization unit may be the group level. For example, if you are disseminating information through community events, you are probably interested in the effect of the intervention on the community. If you expect the intervention to take effect privately, it may be possible to create an intervention aimed at individuals. For example, if an individual receives the intervention while they are on the internet and not aware whether or not their peers have viewed a video, this could be measured at the individual level.

**Logistical feasibility:** Sometimes the preferred unit of randomization is logistically impossible. Consider again our example: If we decide that we want to randomize on the individual level because we are interested in individual effects of our information campaign, and the information is spread through events in community centres, it may not be possible to ensure that the control group does not participate in these events. Therefore, it may not be feasible to randomize on the individual level.

**Political feasibility:** Communities or partners of the project may demand a certain level of randomization because they want to ensure that all people benefit from the project in the long run.

</div>

## *Common pitfalls*

While RCTs are best practice to identify the causal effect of an intervention, there are some common pitfalls which should be considered.

"**Spillovers"** are the term that social scientists use to refer to the fact that sometimes people in the control group will be exposed to the intervention. For example, imagine you are testing the effect of giving a vaccine to children in school. You give the vaccine to the treatment group, but not to the control group. Not only do you lower the rates of illness within the treatment group, you also prevent them from contaminating the control group. This means that the control group is less sick than it would have been had the intervention not been rolled out. If this were to happen, comparing the treatment against the control would give you an artificially low estimate of the impact because the control group no longer represents an appropriate counterfactual.

**Attrition:** "Attrition" refers to when people either stop participating in the intervention or leave the study (so that it becomes impossible to collect their data). This can happen if individuals drop out of the study, refuse to answer questions regarding the outcome, or cannot be found for the end line survey. When people leave the study it complicates the analysis, because it makes the treatment and control group difficult to compare against each other. This typically means that the researcher has strong incentives to ensure that those who do participate in the study stay involve through to the end.

**Types of field experiment**

So far we have described RCTs. As discussed, RCTs use random assignment to separate the subjects into treatment and control groups, allowing the organization to understand the effects of an intervention that is applied only to the treatment group.

Another method for consideration is a lab-in-the-field experiment. Lab-in-the-field experiments combine elements of laboratory experiments and field experiments like the RCTs we described above. This approach places a standardized experimental methodology in a naturally occurring environment and incorporates elements of that environment into the experiment.

# Impact Measurement

Independent of the method used to test an intervention (e.g. a field experiment, a lab-in-the-field approach), accurately measuring the effect of the intervention is critical. In particular, it is often helpful to measure how an intervention affects actual behaviour, especially if the program's intent is to change behaviour and not just attitudes. Measuring the outcome of an intervention is often a challenging task. If an intervention intends to change behaviour, the best way to measure its impact is to observe the targeted behaviour itself rather than relying on self-reported behaviour changes. This can be difficult for two main reasons:

- First, it may be logistically impossible to directly observe the behaviour, for instance because it is illegal (e.g. domestic violence) or simply not observable ex-post (e.g. prenatal sex-selection).
- Second, it might be hard to rely on interview or self-reported data, because the behaviour might be illegal (e.g. wildlife trade) or culturally sensitive (e.g. female genital cutting).

Beyond the standard methods that rely on self-reports, interviews, and focus groups, other methods can be used to measure behaviour. This section discusses two main approaches, measuring in the field or in a lab (and for a lab, there are a few ways to do this).

## *Measuring behaviour in the field*

Surveys or lab experiments can face two major methodological challenges: experimenter demand effects and external validity.

- **Experimenter demand effects** refers to the potential for the experimenter to influence participant behaviour due to beliefs and cues about the result the participant thinks the experimenter wants. For example, a participant may tell a researcher that she would never buy ivory jewellery, even if she might actually do so when not being observed.

- **External validity** refers to the extent to which the results of a study can be generalized to other settings (ecological validity) and other people (population validity). Generally, surveys of representative samples should externalize to the population, but this might not adequately capture the group you intend to study. For example, a study of the local market for tiger skins in one part of India may not be at all applicable to a market in Western Europe.

Measuring behaviour in the field helps to overcome these challenges. Here, we describe several methodologically robust and appropriate ways of doing so, for consideration of adoption by demand reduction practitioners.

### *Online and Administrative Sources*

A considerable amount of illegal wildlife trade is done online. Online and administrative sources are a useful tool to gather information about behaviour and trends in markets. Some new, innovative studies have used online data to detect racial bias, for example, among users of Airbnb or through Google's search trends.

One example of such methods being employed in the wildlife space is an investigation carried out in 2014 by the International Fund for Animal Welfare, which looked at trade patterns in endangered wildlife in 280 online marketplaces across 16 countries.[6] This study identified almost 10,000 advertisements featuring over 30,000 wildlife parts and products from endangered species for sale at an estimated value of over 10 million US Dollars. Advertisements are a good indicator of supply, and contextualise the strength of demand. To identify consumer desire, online traces that sales have been made could be investigated. An econometric methodologist should be engaged to assess the representativeness of the data being collected to ensure it is informative.

Data on convictions and confiscations provides another set of useful contextual information. When an intervention in a specific region is successful and the demand for illegal wildlife decreases, this should correlate with a reduction in the number of convictions for purchasing illegal wildlife products (assuming consistent law enforcement). An advantage of this approach is that the data collection can be relatively low-cost and implemented at a large scale.

---

**Case study: Measuring racism on AirBnb**

One experiment carried out in the United States measured the response of AirBnb hosts to potential guests with distinctly African American sounding names and those with distinctly white sounding names, who were otherwise identical in how they presented to the hosts. It found that the guests with African American sounding names were 16% less likely to be accepted by the Airbnb hosts.

The researchers were able to access data on the hosts, such as age and race, as well as their reviews from previous guests, so they could also see whether they had hosted African American guests in the past. They were also able to gather data on the property and the demographics of the neighbourhoods studied, including racial demographics.

Based on their investigation, the researchers concluded that the type of discrimination they measured is most prevalent among hosts who have never hosted an African American guest, suggesting consistent discrimination over time by a small group of hosts.

The study led to AirBnb reforming their standards in order to address the problem of racial discrimination.

Source: Edelman and Luca (2014)

---

[6] TRAFFIC China resources on relevant issues: https://www.traffic.org/publications/

***Social listening***

One way to monitor demand is through online social listening. Social listening is a form of audience research that enables the researcher to assess consumers' sentiment towards a product by analysing data and conversations from online retail platforms.  It is commonly used for business analytics and making brand strategy by companies who track and analyse how their brand functions online. Online social listening is a useful method to analyse forms of previously overlooked social media engagements (Crawford, 2009) in social science research as well. The process of social listening would require the researcher to:

1. Monitor social media mentions about the product
2. Analyse insights by going beyond the data to assess the style and mood of the mentions (thereby differentiating it from mere social media monitoring)
3. Use a variety of social listening tools such as NetBase to track conversations online about the product and applying natural language processing to get 'true' consumer sentiment about the product.

Online social listening programmes can offer insight about consumers' attitudes towards a wildlife product which may be reflective of the product's demand. Tracking online retail channels for illegal wildlife products would require a nuanced understanding of the product. Then, identifying social mentions of the products would require the knowledge of markets in which the product is sold, and having access to retailers' sales channels on various social media platforms through the course of the intervention.

Many applications of social listening come from its use in business related campaigns or PR campaigns. These include assessments of the sentiment and attitudes of customers. In light of this, social listening may offer potential in e.g. ascertaining the mood of customers towards an illegal wildlife product such as ivory. Users and retailers of ivory can be traced online via mentions of products that use ivory (or equivalent terms). In the event a retailer or retail network is identified, the nature of the individual conversations about ivory related products might be analysed at baseline and at the endline to look for a change, if any, in the nature or sentiment of the conversation that is taking place about ivory products, while being cognisant of the changes in laws, prices and other parameters for the product.

As opposed to social media analysis (e.g. analysing advertisements of ivory to look for a rise/fall in demand), social listening allows for a detailed understanding of the specific nature of customers' demand. The disadvantage of such an approach is that social listening involves access to a lot of personal data ownership issues on part of the researchers, and also in stating precisely, the extent to which the intervention brings about any noticeable attitudinal change.

***Observation***

If the behaviour of interest is observable, one option is to set up a process that enables you to measure it through direct or indirect observation. A common practice in household surveys implemented by institutions such as the World Bank, is to do a review of the household in which the interviewees live. For example, the enumerator might observe some of the main rooms in

the household while conducting the interview to see if the household has particular items. In some cases, it may be possible to look for items that indicate whether or not an intervention was successful.When behaviour is complex or illicit, this can be challenging. It can be done, however. For example, one study measured female genital cutting by observing the henna-decorated feet of young girls, a unique signal that they have been circumcised recently (Vogt et al., 2016). Another study on the quality of medical advice provided in India and Paraguay recruited non-medical personnel to spend a day with the doctors, observing and making notes on all of their interactions with patients throughout the day. The exercise included noting their conversations, examinations, what treatments they prescribed, and how much they charged. The researchers aggregated the information into an 'effort index'. The study found that the quality of care provided is a function of low competence, compounded by low levels of effort on the part of the doctors (Das, Hammer, and Leonard 2008).

The advantage of this method is that it can directly measure *actual* behaviours in real world contexts. The disadvantage is that this can be very difficult to do when behaviours are complex or illicit, and it requires substantial training for implementers, alongside ethical, legal, personal security, and other similar considerations.

### Measuring illicit behaviours using unmatched count techniques
Another survey method commonly used to measure attitudes on sensitive topics is the item count technique or list experiment. This topic was previously addressed in section 6 of *Reducing Demand for Wildlife Products.* In this survey method, people are not directly asked about their attitudes or experiences. Rather, they are presented with a list of statements and asked how many items they answer affirmatively for the list as a whole. For the treatment group, the list contains a sensitive question or statement next to a set of innocuous questions or statements. In the control group, the list contains only the innocuous questions or statements. The difference between the mean of affirmative responses in the treatment and the control group can be used to compute the average approval to the sensitive question/statement (De Cao and Lutz, 2018). List experiments have also been used to estimate the prevalence of illegal activities such as bushmeat hunting in National Parks (e.g. Harrison et al. 2015).

### Collaborating with Governments and international development partners
Governments and international organizations (nonprofits and international institutions) frequently collect household consumption data to estimate the socioeconomic conditions in countries. These surveys are often extremely detailed and cover a wide range of subjects. One strategy would be to collaborate with these institutions to include questions about the consumption of endangered species. The questions could be carefully embedded within consumption modules so as to seem natural. This approach would have two advantages. First, it may enable enumerators to ask these questions without raising suspicion that the purpose of the question is to understand demand for illegal wildlife products. Second, the samples are typically constructed with care to ensure that generalizable claims can be made from the data. Thus, it would be clear how generalizable the findings from such studies would be.

## *Measuring behaviour in mobile lab*

If it is possible to accurately identify the target consumer group and convince them to participate in surveys or lab-in-the-field activities, then it may be possible to test interventions in a more controlled fashion and develop a more granular understanding of whether interventions are working. A mobile lab is a room that is temporarily set up to enable study participants to answer questions and play behavioural games in private (see image below). In a typical behavioural game, study participants might be given money and asked to make decisions about whether to invest it or share it with others.



Credit: Amy Elhaldi
Source: Vogt et al. (2016) and Efferson et al. (2015)

Some key strengths of mobile lab studies are:

1. **They are private.** Study participants engage with the computer or tablet by themselves and do not report their answers to an enumerator. This reduces the risk that the answers participants report are based on concern for social desirability.

2. **There is less measurement error.** The highly controlled environment in mobile labs means that much of the randomness that affects the outcomes measured in the field is reduced or eliminated.

3. **It is sensitive and realistic.** The functionality of the applications on the tables means that it is possible to simulate complex scenarios that invite participants to psychologically transport themselves into the decision-making context of the act and make decisions that have real implications. This activity enables the researcher to develop a much richer and more sensitive understanding of the behaviour.

4. **It is context sensitive:** Mobile labs allow the researcher to explore decision-making within the local context and to have a meaningful, representative subject pool.

5. **It is social.** Unlike early pen and paper studies, mobile labs allow the researcher to look at how people interact with each other in real-time. This means that you can look at whether other people's decision to undertake an act within the context of the study affects another person's propensity to do so.

### *Measuring social norms using incentivized Vignette Studies*

An important correlate of behaviour is the social norm that sustains the behaviour. Social norms are behavioural rules that people prefer to conform to because they believe others in their group either expect or demand it of them (Bicchieri 2006). A common way to measure perceived social norms is through an incentivized vignette. In this method, you can ask participants their own behaviour, and then let them guess about another participants' behaviour (Krupka and Weber 2018). Thus, a businessman in Vietnam might first answer a question about whether he consumes an illegal wildlife product, and then he will guess the answer of another randomly selected businessman in the study. Participants can be incentivized to try to guess correctly by them a reward for every correct answer. This 'incentivized vignette method' is a reliable way to measure perceived social norms or focal points in a well-defined group of people.

### *Measuring implicit attitudes using the implicit association test*

Explicit attitude measurement can suffer from misreporting either because people prefer not to report their genuine attitudes, beliefs, and behaviours or because they are not aware of them. Implicit attitudes are evaluations we make that are beyond our conscious awareness. A popular measure of implicit attitudes is the implicit association test (IAT, Nosek et al., 2007). IATs have been used to measure implicit attitudes toward race, gender, illicit behaviours, and other social phenomena.  IATs investigate the strength of mental associations by measuring response time when participants categorize neutral target stimuli, for instance race, and value stimuli, for instance positive or negative expressions. The IAT is increasingly used in development research to measure the impact of evaluations (Beaman et al., 2009; Vogt et al., 2016). Implicit association tests are most effectively run in lab or lab-in-the-field settings. They can be implemented with the help of a trained behavioural scientist.

### *Measuring illicit preferences using computerized data collection and privacy*

Many surveys, even when done in person, are handled by computer assisted personal interviewing instead of pen and paper methods. Research has shown that, for the most part, this is cheaper and, more importantly, introduces fewer errors and therefore produces higher quality data than simple pen and paper collection. Using tablets, mobile laboratories, or smart phones also allows for more privacy, less threat of social desirability bias, and smoother collection of experimental and behavioural data.

In areas where illiteracy is very high, it is possible to set up data collection in a way that participants will be able to answer some questions completely on their own, using tablets with audio or video instructions. This provides a high level of privacy to participants, which is

especially important when questions are culturally sensitive. The privacy might also encourage honest responses to questions where there is a strong temptation to give the socially desirable answer. Computerized data collection also allows to use more advanced methods to counter social desirability biases, such as the implicit association tests or incentivized vignette methods described above and can randomize questions across subjects. It also allows researchers to collect more behavioural data – participants can be connected to each other and play behavioural games. This can be used to measure, for instance, decision making and bargaining power within couples or extended families.

# Summary of Bangkok Working Group

On November 28ᵗʰ 2018 a Working Group came together during the "Revisiting Efforts to Reduce Demand for Illegal Wildlife Products: Showcasing Best Practice in Behavioral Change" conference, to discuss options to make methodological improvements to learning and evaluating interventions to address wildlife demand. The insights and proposals in this Options Paper were presented and discussed among the participants. Through the discussion, three main recommendations were made.

The recommendations are summarized below alongside an associated next step. In addition, practical resources on empirical methods are provided below.

**Recommendation 1:** Evaluation should be baked into intervention design *from the beginning* to enable program leaders to implement their program in ways that facilitate learning for future projects. As a first step, all projects that have a learning and evaluation component should have a baseline. More innovative projects should endeavour to randomize the treatment assignment to enable the conservation field to understand which strategies are working and which strategies are not.

*Suggested next step:* *Identify partner policy researchers to provide technical assistance in program design.*

**Recommendation 2:** New approaches to measurement should be developed to advance the community of practice's understanding of trends in demand and the effectiveness of demand reduction efforts. Data that comes from self-reported measures may not be reliable. Behavioural and observational measures to data collection were noted as a promising new way to measure demand behaviours.

*Suggested next step:* *Set up a joint task force of practitioners and policy researchers to trial innovative methods to measure illicit wildlife consumption behaviors.*

**Recommendation 2:** Data should be kept source open to enable teams to build shared approaches to measurement. Evaluations should be frequent and build upon one another rather than be treated as one-off events.

*Suggested next step:* *Include empirical methods as a theme within the community of practice led by TRAFFIC.*

Resources:

| Document | Author |
|---|---|
| *Running Randomized* Evaluation*s: A Practical Guide* | Kudzai Takavarasha and Rachel Glennerster |
| *Test Learn Adapt* | Behavioural Insights Team |
| *Monitoring and evaluating behaviour change amongst illegal wildlife product consumers good practice guidelines for Social and behavioural change communications practitioners and communications professionals* | Gayle Burgess (TRAFFIC) |

# Next Steps

In light of the consumer-focused methodological options summarised here and discussed at the Working Group in Bangkok, November, 2018, TRAFFIC will prepare a larger proposal for a demand reduction impact measurement methodology that incorporates these alongside other more objectively verifiable measures. The initial output of this work will be a draft paper summarising the methodology, for discussion amongst an external expert group.

Further information regarding this work, is available from Gayle Burgess, in the first instance:

**Gayle Burgess**
**TRAFFIC | Behavioural Change Coordinator**
David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK
Email: gayle.burgess@traffic.org |  Tel: +44 7392 197748
Website | Facebook | Twitter | LinkedIn | Sign up to our newsletter
UK Registered Charity No. 1076722, Registered Limited Company No. 3785518

# References

Angrist, J.D. and Pischke, J.S., 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124 (4): 1497-1515.

Bechtel, Robert B., and John Zeisel. 1987. Observation: The world under a glass." In *Methods in environmental and behavioural research*, edited by Robert B. Bechtel, Robert W. Marans, and Wiliam Michelson, 11–40. New York, NY: Van Nostrand Reinhold

Becker, Gary Stanley. 1976. *The Economic Approach to Human Behaviour*. Chicaco: University of Chicago Press.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

Burgess, Gayle, Zain, Sabri, Mulner-Gulland E.J., Eisingerich B. Andreas, Sharif, Vian, Ibbett, Harriet, Olmedo Castro, Alegria and Heather Sohl. 2018. *REDUCING DEMAND FOR ILLEGAL WILDLIFE PRODUCTS*.
available at:
https://www.traffic.org/site/assets/files/11081/demand_reduction_research_report.pdf

Camerer, Colin .F. and George Loewenstein. 2004. "Behavioural Economics: Past, Present, Future." In *Advances in Behavioural Economics* edited by Camerer, Colin .F, Loewenstein, George and Matthew Rabin, 3–51. Princeton: Princeton UniversityPress.

Chetty, Ray, Friedman, John N., Leth-Petersen, Soren, Nielsen, Torben Heien and Tore Olsen. 2014. "Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark." *Quarterly Journal of Economics* 129: 1141–1219.

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum*, 23(4), 525-535. doi: 10.1080/10304310903003270

Das, J., Holla, A., Das, V., Mohanan, M., Tabak, D. and Chan, B., 2012. In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. *Health affairs*, *31*(12), pp.2774-2784.

De Cao, Elisabetta and Clemens, Lutz. 2018. "Sensitive survey questions: Measuring attitudes regarding female genital cutting through a list experiment." *Oxford Bulletin of Economics and*

*Statistics* 80(5): 871–892.

Edelman, Benjamin G., and Michael Luca. "Digital discrimination: The case of Airbnb. com." (2014).

Flores, A., & Bachofen, A. Social Listening: 3 Case Studies | News for Authors. Retrieved from http://authornews.penguinrandomhouse.com/social-listening-3-case-studies/

International Fund for Animal Welfare. 2014. *Wanted – Dead or Alive.*
available at: https://www.ifaw.org/european-union/resource-centre/wanted-dead-or-alive-exposing-online-wildlife-trade

Krupka, Erin L. and Roberto A., Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?." *Journal of the European Economic Association* 11(3): 495–524.
https://doi.org/10.1111/jeea.12006

Lira, Andreas Gonzales and Ahmed Mushfiq Mobarak. 2018. "Enforcing Regulation under Illicit Adaptation." COWLES FOUNDATION DISCUSSION PAPER NO. 2143
https://arxiv.org/abs/1808.09887

LONDON CONFERENCE ON THE ILLEGAL WILDLIFE TRADE. 2014.
available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/281289/london-wildlife-conference-declaration-140213.pdf

Olmedo, Alegria, Sharif, Vian and E.J. Milner-Gulland. 2017. "Evaluating the Design of Behaviour Change nterventions: A Case Study of Rhino Horn in Vietnam." Conserv. Lett.
https://doi.org/10.1111/conl.12365.

Nuwer, R. (2018). How Japan Undermines Efforts to Stop the Illegal Ivory Trade. Retrieved from https://www.nationalgeographic.com/animals/2018/09/japan-illegal-ivory-trade-african-elephants/

Roberton, Scott I. 2014. "Has Demand for Rhino Horn Truly Dropped in Vietnam?" *Wildlife Conservation Society*
available at: *http://voices.nationalgeographic.com/2014/11/03/has-demand-for-rhino-horn-truly-dropped-in-vietnam/*

TRAFFIC. 2016. *Changing behaviour to reduce demand for illegal wildlife products: workshop proceedings.* TRAFFIC, Cambridge, UK.
available at:
http://www.trafficj.org/publication/16_Changing_Behaviour_to_Reduce_Demand_for_Illegal.pdf

TRAFFIC. 2016. *SBCC Good Practice Guidelines.* TRAFFIC, Cambridge, UK. available at: https://www.traffic.org/publications/reports/sbcc-good-practice-guidelines/

Veríssimo, Diogo., Schmid, Carina,, Kimario Fidelcastor F and Heather. E. Eves. 2018. "Measuring the impact of an entertainment-education intervention to reduce demand for bushmeat." *Animal Conservation* 21: 324–331.

Vogt, Sonja, Zaid, Nadja Ahmed Mohmmed, Ahmed, Hilal El Fadil, Fehr, Ernst and Charles Efferson. 2016 "Changing cultural attitudes towards female genital cutting." *Nature.* 538 (7626): 506–509.

World Bank. 2015. World Development Report 2015: *Mind, Society, and Behaviour.*