Opinion paper

The bucket and the searchlight: formulating and testing risk hypotheses about the weediness and invasiveness potential of transgenic crops

Alan RAYBOULD*

Syngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, UK

The bucket and the searchlight are metaphors for opposing theories of the growth of scientific knowledge. The bucket theory proposes that knowledge is gained by observing the world without preconceptions, and that knowledge emerges from the accumulation of observations that support a hypothesis. There are many problems with this theory, the most serious of which is that it does not appear to offer a means to distinguish between the many hypotheses that could explain a particular set of observations. The searchlight theory proposes that preconceptions are unavoidable and that knowledge advances through the improvement of our preconceptions – our hypotheses – by continuous criticism and revision. A hypothesis is a searchlight that illuminates observations that test the hypothesis and reveal its flaws, and knowledge thereby increases through the elimination of false hypotheses. Research into the risks posed by the cultivation of transgenic crops often appears to apply the bucket theory; many data are produced, but knowledge of risk is not advanced. Application of the searchlight theory, whereby risk assessments test hypotheses that transgenic crops will not be harmful, seems to offer a better way to characterise risk. The effectiveness of an environmental risk assessment should not be measured by the size of the bucket of observations on a transgenic crop, but by the power of the risk hypothesis searchlights to clarify the risks that may arise from cultivation of that crop. These points are illustrated by examples of hypotheses that could be tested to assess the risks from transgenic crops and their hybrids becoming weeds or invading non-agricultural habitats.

Keywords: risk assessment / problem formulation / data requirements / scientific method / risk characterisation

INTRODUCTION

In an essay published in German in 1949 and in English in 1972, Karl Popper described two theories of scientific knowledge, which he called the bucket and the searchlight (Popper, 1979). In the bucket theory, knowledge is believed to originate from observations that are collected without preconceptions. These observations accumulate in a metaphorical bucket and accrete into knowledge. In the bucket theory, observations come first, and hypotheses arise by generalisation from those particular observations – a method called induction. Eventually there may be sufficient observations supporting a hypothesis that it becomes regarded as a statement of a general truth about nature.

An alternative theory is that knowledge increases through observations that test our preconceptions, and it is by observing that our preconceptions are wrong that we gain knowledge. In this theory, observations are made in response to preconceptions; that is, we always have expectations or hypotheses that guide our observations. Hypotheses are searchlights that illuminate our observations and give them their significance. In the searchlight theory, we use general hypotheses to deduce particular expected facts, and when our observations differ from what we expected, we formulate new hypotheses in attempts to eliminate the flaws that led to the erroneous expectations. Knowledge thereby grows by repeated testing and correction of hypotheses.

Induction has proved problematic as a logical foundation for science; for example, however many facts are added to the bucket, it is never possible to prove that no subsequent observation will contradict the generalisations drawn from those facts. A second problem is that infinite generalisations can be drawn from any set of observations, and simply adding similar observations to the bucket does not help to discriminate among those

---

* Corresponding author: alan.raybould@syngenta.com

generalisations. Popper (1959) offered deduction as a solution to the problem of the logic of science. He proposed that we discriminate between hypotheses by searching for experimental conditions under which the hypotheses make different predictions. Hypotheses that make accurate predictions are corroborated and survive for further testing, whereas hypotheses that make inaccurate predictions are revised or discarded. Popper argued that it is active criticism of hypotheses, not the accumulation of facts in favour of hypotheses, which advances science. In Popper's view, hypotheses are searchlights that show which observations are valuable; they are not generalisations that emerge from a bucket of facts.

This paper examines environmental risk assessment for cultivation of transgenic crops in terms of the bucket and searchlight theories. It considers why extensive data on the environmental effects of cultivating transgenic crops often seem to confuse rather than clarify decision-making about cultivation approvals for transgenic crops. It discusses how hypotheses may be used as searchlights to illuminate the most effective data for characterising the environmental risks from transgenic crops, and in particular discusses the relationships between three crucial elements of risk assessment: conceptual models, risk hypotheses and statistical tests. Assessment of the weediness and invasiveness potential of transgenic crops is used to illustrate these ideas.

ENVIRONMENTAL RISK ASSESSMENT FOR TRANSGENIC CROPS – BUCKET OR SEARCHLIGHT?

Over the past 20 years, there has been extensive research into the environmental risks of cultivating transgenic crops (*e.g.*, Brookes and Barfoot, 2008; Economidis et al., 2010; Kessler and Economidis, 2001; Sanvido et al., 2007); however, many of the results seem to have little relevance to regulatory decision-making (Craig et al., 2008; Raybould, 2007), and may have increased rather than allayed public unease about the application of biotechnology to agriculture (Johnson et al., 2007). Given that risk is the product of the likelihood and seriousness of harm that may arise from an activity (*e.g.*, Wolt et al., 2010), two aspects of much of this research are striking. First, there is often little or no attempt by researchers characterise risk; the research simply analyses effects without specifying whether they are harmful. A second related characteristic is that explicit hypotheses about how harm may arise are not tested. Statistical null hypotheses of no difference between a transgenic crop and a comparator may be tested (*e.g.*, Marvier et al., 2007; Squire et al., 2003), but the results of those tests are not evaluated in terms of hypotheses about risk (Raybould, 2007).

Without definitions of harmful effects, and hypotheses about how they may arise, environmental risk assessment research on transgenic plants appears to be an attempt to apply the bucket theory of scientific knowledge. Observations of the environmental effects of transgenic plants are collected and somehow a characterisation of risk that is useful for decision-making is supposed to emerge. This is an example of the "deficit model" of policy making (Lawton, 2007), whereby poor policy is imagined to result from a deficit of facts. The deficit model implies that policy can be formulated directly from the results of research; however, this is not possible, because while science can make predictions about the consequences of certain policies, it cannot determine whether those consequences are desirable (Lubechenko, 1998). Just as knowledge cannot emerge from a bucket of observations, neither can policy.

It is easy to see how the bucket theory can lead to research that may hinder rather than help decision-making and increase unease about the cultivation of transgenic crops. If characterisation of risk is supposed to emerge from observations, then if one is uncertain about the level of risk, the solution is to collect more observations to add to the bucket. The observations cannot solve the problem of what to consider as harmful, but they appear important because they have been collected for the stated purpose of risk assessment. As more data accumulate, decision-making appears more complicated: we believe that the data must be relevant – they are the results of risk assessment research – but we are no nearer to characterising risk unless harm is defined. So, research has been done, but the characterisation of risk has not improved, leading to the understandable conclusion that the research revealed something of concern (Johnson et al., 2007).

The searchlight theory offers a better method for characterising risk. The first stage is to define what would be harmful effects of cultivating the transgenic crop. Laws and non-statutory regulations set out broad objectives of environmental policy called protection (or management) goals. From these objectives, the risk manager and risk assessor agree specific targets for protection; these targets are called assessment endpoints, and comprise an entity, such as a particular species, and a property of that entity, such as its population size in a given area (Suter, 1990). Harm may then be defined as a specified reduction in the abundance, quality or other measure of the condition of the assessment endpoints.

The next stage is to formulate testable hypotheses that guide the collection of observations that allow characterisation risk. The formulation of these hypotheses is considered in more detail below, but in general they are propositions that a link in a chain of events that lead from cultivation of the transgenic crop to a harmful effect is absent; for instance, if harm could arise as a consequence

of hybridisation between a transgenic crop and a species of wild plant, a suitable hypothesis would be that no hybrids will form in places where they could have adverse effects on the assessment endpoints (Raybould and Cooper, 2005; Raybould and Wilkinson, 2005). These hypotheses are searchlights to illuminate the data that are necessary to characterise risk, which in this case could be laboratory data on hybridisation between the crop and the wild relative, or the results of surveys of whether the crop and wild relative grow together near protected habitats. If the hypotheses are corroborated after rigorous testing, one may have confidence that the risks from cultivating the transgenic crop are low and that no further data are required to make a decision. If the hypotheses are falsified, one may conclude that the risks are high, and again decide that no further data are required, or one may decide to test new hypotheses about whether the crop may cause harm.

The advantage of the searchlight method is that it identifies those observations that can characterise risk most effectively. First, data are not accumulated with the idea that they can indicate what is harmful. Secondly, data are obtained under conditions that provide rigorous tests of hypotheses; that is, circumstances that are designed to reveal errors in the hypotheses that the transgenic crop will cause no harm[1]. This targeted approach should minimise the amount of data consigned to the bucket of unknown relevance for risk assessment.

CONCEPTUAL MODELS, RISK HYPOTHESES AND STATISTICAL TESTS

Derivation of risk hypotheses from a conceptual model

If the searchlight theory is correct, risk assessment must be more than a collection of observations on the effects of transgenic crops from which conclusions about risk somehow become apparent. We need testable hypotheses to guide our observations and help us to characterise risk. The first step in the production of such testable hypotheses is a conceptual model, scenario or pathway that describes how cultivation of the transgenic plant may cause harm. At this stage, the links in the chain of events from cultivation to harmful effects are logical: what are the

---

[1] Throughout this paper, risk hypothesis are described as hypotheses that no harm will occur as a result of cultivating a transgenic crop. Strictly, decision-makers will determine whether it is sufficiently likely that no *unacceptable* harm will result. Consequently, a risk assessment could be regarded as a test of the hypothesis that there is a sufficiently high probability of no unacceptable harm; however, for convenience and conservatism, here the assessment is described as a test of the hypothesis of no harm.

necessary conditions for to harm arise, not what is the likelihood of those conditions occurring.

There may be infinite ways by which harm could arise, with the creation of logically consistent scenarios being limited only by our imagination. It is necessary, therefore, to reduce the number of scenarios that will be used to generate hypotheses for testing in the risk assessment. Some logically possible scenarios may appear so implausible that it is almost inconceivable that they pose any risk, and therefore they are not evaluated in the risk assessment. This winnowing out of implausible scenarios may be unconscious or implicit, because some scenarios are so far-fetched that they are not actually thought about, or it may be explicit in that the scenario is considered and rejected.

It is interesting to examine what happens when a scenario is rejected as too implausible to require further consideration. We may propose that cultivation of the transgenic crop could lead to event A which leads to event B which leads to event C that would be considered harmful. This scenario is implausible if at least one of the hypotheses that the chain is incomplete is rigorously tested and corroborated by existing observations: cultivation does not lead to event A; or event A does not lead to event B; or event B does not lead to event C. If we do not make explicit the hypotheses and the tests that corroborate them, it could appear that deeming a scenario implausible is an unjustified assumption; therefore, for all implausible scenarios, one should be able to argue that a condition necessary for harm to occur *via* that pathway is highly improbable based on existing knowledge.

The same method of hypothesis formulation is used to evaluate the plausible scenarios within the risk assessment; however, within the risk assessment the hypotheses and their testing should be explicit. The plausible scenarios are examined in terms of discrete steps that must occur for the cultivation of the transgenic crop to result in harmful effects. From each step it is possible to formulate a hypothesis, which if corroborated or falsified by suitable testing, would characterise risk in a form that is useful to decision-makers. Hypotheses could take several forms: event A does not lead to event B; event A leads to event B at a frequency below that which would cause harm; or event A leads to event B, but event B is below the magnitude necessary for harm (Raybould, 2006, 2010). In each case, the hypothesis can be regarded as a hypothesis of no harm from cultivation of the transgenic crop. Testing hypotheses of no harm, with new studies, with existing data collected for purposes other than the current risk assessment, or both, is the basis of risk characterisation.

The hypotheses of no harm have recently been described as "risk hypotheses" (Raybould, 2006, 2007). The term risk hypotheses had been used previously

to describe the conceptual model (*e.g.*, Patton, 1998); however, this use seems problematic in a scientific context. The conceptual model describes logical relationships between phenomena: *if* cultivation of the transgenic crop leads to event A, *and if* event A leads to event B, *and if* event B leads to harmful event C, *then* cultivation of the transgenic crop will cause harm. The conceptual model is a hypothesis in the logical sense that it is a statement assumed to be true for the sake of argument, but not a hypothesis in the empirical sense that it is a refutable conjecture about the world. In effect, the conceptual model is a logical device that enables the formulation of testable hypotheses of no harm. As it is these testable hypotheses that are the searchlights that illuminate observations for risk characterisation, the term "risk hypothesis" is used in that sense in this paper.

The conceptual model is a pathway that would lead to harm should all the steps be realised. Consequently, a second problem with calling the conceptual model a risk hypothesis is that it suggests that the general hypothesis under test in the risk assessment should be that cultivation of the transgenic crop will be harmful; in other words, there exists a harmful effect of cultivating the transgenic crop. This formulation cannot be correct under the searchlight theory, because it is an example of a strictly *existential* statement – a statement which asserts that something exists – and such statements are irrefutable (Popper, 1959). Refutable hypotheses are strictly *universal* statements – statements which assert that something does not exist. This point can be illustrated with reference to the colour of swans[2]. The strictly existential statement that there is a black (*i.e.*, non-white) swan cannot be refuted, no matter how many white swans are observed. On the other hand, the strictly universal statement that all swans are white, in other words that there are no black swans, can be refuted by the observation of a single black swan; thus, we can refute the hypothesis of no harm – "there are no black swans" – but we cannot refute a hypothesis of harm – "there is a black swan". Hence, if we consider a risk assessment to follow the searchlight theory of scientific knowledge, and if we consider the cultivation of transgenic crops not to be harmful *in principle*, risk hypotheses must be hypotheses of no harm. It follows that low risk is indicated by corroboration of hypotheses of no harm, not by falsification of hypotheses of harm. The alternatives are to regard cultivation of transgenic crops as harmful in principle, in which case we do not need a risk assessment, or to collect bucketfuls of white swans in the impossible attempt to falsify the hypothesis that there is a black swan.

---

[2] In the Logic of Scientific Discovery, Popper discusses statements about black and white ravens; however, statements about white and black swans are used more commonly to illustrate Popper's ideas.

A final point about the formulation of risk hypotheses is that risk assessment is not scientific research and does not create scientific knowledge for its own sake (Hill and Sendashonga, 2003). It organises existing information, along with sufficient new observations, so that a decision can be made. This means that a risk hypothesis should not attempt to make predictions with more precision than is necessary to make a decision (Raybould, 2007, 2010). In the case of hybridisation between a transgenic crop and a wild relative, it may be sufficient for decision-making to test the hypothesis that no hybrids will form in a given area. If that hypothesis is sufficiently precise, it is not necessary to create a model that attempts to predict exactly how many hybrids will form. Precise quantitative predictions are only valuable for risk assessment if decision-making criteria are similarly precisely quantitative.

Testing risk hypotheses

Corroboration or refutation of a risk hypothesis may depend upon the results of statistical tests; however, that does not imply that the characterisation of risk can be reduced to the significance of a statistical test (Suter, 1996). Say, for example, one was assessing the environmental risk from cultivating a variety of transgenic cotton. One of the risk hypotheses may be that transformation has introduced no harmful unintended changes into the cotton. One test of that hypothesis could be a comparison of gossypol concentrations in the transgenic cotton and a near-isogenic line because increased amounts of gossypol could harm non-pest species that feed on cotton tissue. Gossypol concentrations would be measured in the transgenic and non-transgenic lines under controlled conditions, and a suitable test applied to determine the statistical significance of any difference in the average gossypol concentration between the lines. Lack of statistical significance would corroborate the risk hypothesis of no harmful unintended changes, but alone would not provide sufficient corroboration for decision-making unless gossypol concentrations were the only indicator of potentially harmful unintended changes.

A statistically significant test would falsify the null hypothesis of no difference in gossypol concentrations between the transgenic and non-transgenic lines, but alone it would be insufficient to falsify the risk hypotheses because the difference may be insufficient to produce a harmful effect. To evaluate the relevance of the finding, one could compare the concentration of gossypol in the transgenic line with concentrations found in currently cultivated cotton varieties, and finding that concentrations in the transgenic variety were within the range of most other varieties would corroborate the risk hypothesis. If the comparison indicated an unusually high amount of gossypol in the transgenic variety, the effects of that

concentration on representative non-pest species could be tested in the laboratory; and if those effects were adverse, one could use semi-field or field studies to see whether those adverse effects were realised in the field.

This example indicates a general concept for testing risk hypotheses, which is that initial tests are made under conservative conditions designed to minimise false negatives. Returning to the example of swans, initial tests seek to minimise the possibility that indicators of harmful black swans are mistakenly identified as indicators of harmless white swans. Minimisation of false negatives means that if tests produce results which suggest that harm is unlikely – only white swans are indicated – one has strongly corroborated the risk hypothesis and may decide it requires no further testing. Minimisation of false negatives means that many false positives may be detected; our measures indicate the possibility of black swans when further evaluation would show that they actually indicate white swans. In the gossypol example, a statistically significant difference in gossypol concentration between the transgenic variety and its near-isogenic line is an indicator of a black swan. This difference may later be found to be an indicator of a white swan if gossypol concentrations in the transgenic variety are well within the historical range of gossypol concentrations of other varieties, or if they have no adverse effects on representative non-pest species in the laboratory.

The concept of starting with conservative tests most likely to reveal indictors of harm and only moving to more realistic tests if those indicators are detected is called tiered testing (Garcia-Alonso et al., 2006; Touart and Maciorowski, 1997). It is an effective way to quickly categorise activities into those that pose low risk and require little or no further testing, and those that may pose high risk and require further evaluation if lower risk is to be demonstrated. The criterion for deciding whether further testing is required is a judgement about the best balance between the costs of over-testing some activities that pose low risk and the costs of incorrectly determining that high risk activities pose low risk (Caley et al., 2006; Chapman et al., 1998).

IMPROVING ENVIRONMENTAL RISK ASSESSMENTS FOR TRANSGENIC CROPS

A comparison of the bucket and searchlight theories of knowledge offers several pointers to good environmental risk assessment for transgenic crops:

- *Define* what would be regarded as harmful effects of cultivating transgenic crops in terms of adverse changes to assessment endpoints. If this proves difficult, discuss the aims of policy with risk managers and decision-makers; do not collect more observations in an attempt to *discover* what is harmful.

- Produce conceptual models or scenarios in the form of a series of events that are necessary for cultivation of the transgenic crop to lead to adverse changes to the assessment endpoints.
- From the conceptual models, formulate hypotheses which postulate that events necessary for adverse effects to the assessment endpoints do not occur; these are risk hypotheses of no harm.
- Test the risk hypotheses under conditions most likely to falsify them; in other words, conditions most likely to indicate the potential occurrence of the conditions necessary for adverse effects on assessment endpoints.
- If the potential for harm is not detected, testing under more realistic conditions is unnecessary; if the potential for harm is detected, there is the option to evaluate that potential under more realistic conditions.
- Confidence that cultivation of the transgenic crop poses low risk is established by sufficient corroboration of risk hypotheses of no harm; it is not established by attempts to falsify hypotheses of harm.
- Sufficient corroboration is a judgement that the risk hypothesis has withstood falsification by enough tests of adequate rigour; it is not definitively established by the result of a statistical test.

Application of these ideas to risks posed by weediness and invasiveness of transgenic crops is considered below.

WEEDINESS AND INVASIVENESS ASSESSMENTS FOR TRANSGENIC CROPS

One of the oldest concerns about the cultivation of transgenic crops is that they will become weeds of agriculture or invasive of non-agricultural land (Keeler, 1989). Interest in weediness and invasiveness potential is currently increasing as transgenic crops with abiotic stress tolerance enter regulatory systems (Nickson, 2008; Warwick et al., 2009; Wilkinson and Tepfer, 2009). A complete framework for environmental risk assessment of the weediness and invasiveness potential of transgenic crops is beyond the scope of this paper, and therefore discussion concentrates on how to make the most effective use of data; in other words, to maximise the power of the searchlights.

Defining harmful effects

Legislation relating to the cultivation of transgenic crops often seeks to protect agricultural production and the environment. In the following discussion, the United States regulatory system is illustrated as it is relatively easy to identify its protection goals. Transgenic crops are treated

Table 1. A generic scenario for harm through reduced crop quality resulting from dispersal of transgenic pollen to other crops.

Scenario	Risk hypothesis
Transgenic crop produces pollen	Transgenic crop does not produce pollen
Pollen disperses to neighbouring crops	Pollen does not disperse to neighbouring crops
Pollen fertilises neighbouring crops	Pollen does not fertilise neighbouring crops
Transgenic protein is produced in seed	Transgenic protein is not produced in seed
Transgenic protein reduces crop quality	Transgenic protein does not reduce crop quality

as "regulated articles" by the United States Department of Agriculture (USDA) under the Federal Plant Protection Act (FPPA) (McHughen and Smyth, 2008; National Research Council, 2002). Regulated articles are regarded as potential plant pests and the USDA must grant non-regulated status before transgenic crops can be grown commercially without restrictions on the area planted and movement between states. If a transgenic crop produces pesticides, such as *Bt* proteins, the United States Environmental Protection Agency (US EPA) must also register the pesticide under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) (McHughen and Smyth, 2008) before seeds of the crop can be freely sold.

Regulation of transgenic plants under the FPPA, falls to the USDA's Animal and Plant Health Inspection Service (APHIS), which states is purpose as "safeguard[ing] agriculture and natural resources from the risks associated with the entry, establishment, or spread of animal and plant pests and noxious weeds", and thereby "ensures an abundant, high quality, and varied food supply, [. . .] and contributes to the preservation of the global environment". Thus, the protection goals for environmental risk assessment of transgenic crops may be extremely broad: safeguarding the supply of abundant, high quality and varied food, and preservation of the global environment. The protection goals of FIFRA, "protect[ing] the public health and environment from the misuse of pesticides by regulating the labelling and registration of pesticides and by considering the costs and benefits of their use" are similarly broad. The US EPA must also take account of the Endangered Species Act, which prohibits any action that can adversely affect an endangered or threatened species or its habitat. In compliance with this law, EPA must ensure that use of the pesticides it registers will not harm these species.

For environmental risk assessment of transgenic crops, derivation of assessment endpoints from the protection goals tends to rely on precedent established by previous petitions for non-regulated status and registrations of pesticidal compounds in transgenic plants, and discussions in the literature (*e.g.*, Pimentel et al., 2001). Assessment endpoints are commonly the abundance of particular organisms, especially species that provide valuable functions, such as biological control and pollination, and species that are endangered. Crop yield and quality could also be made operational assessment endpoints (Raybould, 2005); however, it is often easier and more conservative to assume that there is always a negative relationship between weed abundance and crop yield and quality, and to use the abundance of volunteer weeds as the assessment endpoint.

Conceptual models and risk hypotheses

Adverse effects to the assessment endpoints listed above could occur by several pathways. A commonly considered scenario is the expression of transgenic insecticidal proteins that cause harm through adverse ecotoxicological effects on populations of exposed non-target organisms. These scenarios have been considered elsewhere (*e.g.*, Garcia-Alonso et al., 2006; Raybould et al., 2007, 2011; Romeis et al., 2008) and are not considered further in this paper. A second set of pathways involve the dispersal of transgenes from the transgenic crop in seed or pollen and their establishment in new plants through germination of transgenic crop seed or by fertilisation of other crops or wild species by transgenic pollen. These scenarios may all be considered as aspects of weediness or invasiveness potential of the transgenic crop as they involve movement of transgenes to places where they could cause harm.

Generic conceptual models or scenarios are illustrated in Tables 1–5: scenarios 1 and 2 illustrate harm through reduced crop quality resulting from crop-to-crop gene flow by pollen and seed; scenario 3 illustrates harm through reduced crop yield resulting from spatial or temporal dispersal of transgenic seed; and scenarios 4 and 5 illustrate harm through reduced abundance of valued species resulting from dispersal of transgenic pollen and seed to non-agricultural habitats. Each scenario is intended to illustrate a general pathway and more or fewer steps from cultivation of the transgenic crop to the harmful effect could be considered.

The purpose of setting out detailed scenarios is the formulation of risk hypothesis. For clarity, the scenarios have been formulated to generate the most conservative risk hypotheses: the complete absence of particular phenomena. In reality, risk hypotheses may postulate that

Table 2. A generic scenario for harm through reduced crop quality resulting from spatial or temporal dispersal of transgenic seed in agricultural habitats.

Scenario	Risk hypothesis
Transgenic crop produces seed	Transgenic crop does not produce seed
Seed disperses to neighbouring crops/Seed disperses to subsequent crops	Seed does not disperse to neighbouring crops/Seed does not disperse to subsequent crops
Seed germinates in neighbouring crops/Seed germinates in subsequent crops	Seed does not germinate in neighbouring crops/Seed does not germinate in subsequent crops
Transgenic crop establishes	Transgenic crop does not establish
Transgenic crop produces transgenic protein	Transgenic crop does not produce transgenic protein
Transgenic protein reduces crop quality	Transgenic protein does not reduce crop quality

Table 3. A generic scenario for harm through reduced crop yield resulting from spatial or temporal dispersal of transgenic seed in agricultural habitats.

Scenario	Risk hypothesis
Transgenic crop produces seed	Transgenic crop does not produce seed
Seed disperses to neighbouring crops/Seed disperses to subsequent crops	Seed does not disperse to neighbouring crops/Seed does not disperse to subsequent crops
Seed germinates in neighbouring crops/Seed germinates in subsequent crops	Seed does not germinate in neighbouring crops/Seed does not germinate in subsequent crops
Transgenic crop establishes	Transgenic crop does not establish
Transgenic crop affects growth of the crop	Transgenic crop does not affect growth of the crop
Transgenic crop reduces crop yield relative to effects of non-transgenic counterpart	Transgenic crop does not reduce crop yield relative to effects of non-transgenic counterpart

Table 4. A generic scenario for harm through reduced abundance of valued species resulting from dispersal of transgenic pollen to wild species.

Scenario	Risk hypothesis
Transgenic crop produces pollen	Transgenic crop does not produce pollen
Pollen disperses to wild relative	Pollen does not disperse to wild relative
Pollen fertilises wild relative	Pollen does not fertilise wild relative
Transgene is stably introgressed	Transgene is not stably introgressed
Wild relative produces transgenic protein	Wild relative does not produce transgenic protein
The transgenic protein increases resistance to a stressor	The transgenic protein does not increase resistance a stressor
Increased resistance leads to higher seed production	Increased resistance does not lead to higher seed production
Higher seed production increases abundance of the wild relative	Higher seed production does not increase abundance of the wild relative
Increased abundance reduces abundance of valued species	Increased abundance does not reduce abundance of valued species

phenomena occur below a frequency or magnitude necessary for harm. In scenario 1, for example, hypotheses that postulate cross-fertilisation frequencies below a threshold of concern may be more suitable than a hypothesis of the complete absence of gene flow. Nevertheless, as a general rule, starting conservatively is a good tactic because greater conservatism offers more confidence in conclusions of low risk should the risk hypotheses be corroborated (Raybould, 2006).

Testing risk hypotheses

In many cases, the risk hypotheses about weediness and invasiveness potential will be that the weediness or invasiveness potential of a particular transgenic crop is no greater than that of the crop in general. A conservative test of that risk hypothesis is a comparison of the transgenic crop with a non-transgenic near-isogenic line or other suitable comparator for characters that predict

Table 5. A generic scenario for harm through reduced abundance of valued species resulting from dispersal of transgenic seed in non-agricultural habitats.

Scenario	Risk hypothesis
Transgenic crop produces seed	Transgenic crop does not produce seed
Seed disperses to non-agricultural habitats	Seed does not disperse to non-agricultural habitats
Transgenic crop establishes in non-agricultural habitat	Transgenic crop does not establish in non-agricultural habitat
Transgenic crop forms self-sustaining population	Transgenic crop does not form self-sustaining population
Population increases in abundance	Population does not increase in abundance
Increased abundance harms a valued species	Increased abundance does not harm a valued species

weediness or invasiveness potential, evaluated by a statistical test of the null hypothesis of no difference (*e.g.*, Raybould et al., 2010). For extra conservatism, the characterisation should be done under a range of conditions, some of which are most likely to reveal differences; for example, transgenic drought-tolerant maize should be compared with non-transgenic maize under conditions where water is and is not limiting plant growth (Nickson, 2008). Corroboration of the statistical null hypothesis indicates no detectable increase in weediness or invasiveness potential, whereas falsification of the null hypothesis only indicates an effect that has the *potential* to lead to harm; statistical significance *per se* does not falsify a risk hypothesis of no harm.

If statistical differences in characters associated with weediness or invasiveness potential are found, the differences should be evaluated. In effect, this evaluation is a Tier 2 assessment triggered by the finding of statistically significant differences at Tier 1. Higher tier evaluations could take many forms depending on the nature of the difference detected and the route by which that difference could lead to harm. If the difference is unexpected and not an intended effect of transformation, the most suitable Tier 2 test may be a comparison with a wider range of crop genotypes than the near-isogenic line used at Tier 1. If the difference is an intended effect of the transgene, then exploration of the consequences of that difference may be necessary, particularly if the intended phenotype is outside the normal range of the crop.

One approach to evaluating intended differences is to simulate predicted effects of the transgene. If the transgene confers disease resistance, and the pathway to harm involves introgression of the gene into a wild relative, a comparison of seed output between artificially infected and mock-inoculated individuals of the wild relative under controlled conditions may be informative. In this case the mock-inoculated individuals simulate the effect of the disease resistance gene in the wild relative. The outcome of this experiment is not inevitably higher seed output in the mock-inoculated group because the wild relative may be naturally immune to the pathogen, or may tolerate infection and show no reduction in seed production

(Maule et al., 2007; Raybould and Cooper, 2005). In the case of introgression of insect resistance transgenes, the wild relative of the crop could be protected from insect attack with insecticides to simulate the effect of transgenic resistance (*e.g.*, Raybould et al., 1999; Sutherland and Poppy, 2005). Another approach is to assume that the transgene leads to increased seed production and then test the hypothesis that increased seed production does not lead to increased abundance of the wild relative, in other words that the population growth rate is not seed-limited. This test can be performed by sowing seeds at densities greater than those observed in the field, or by excluding seed predators, and observing the number of plants that reach maturity (*e.g.*, Bergelson, 1994; Cummings and Alexander, 2002; Raybould and Cooper, 2005). If transgenic crop × wild relative hybrids are available, they could be used in experiments to evaluate their potential for increased invasiveness (*e.g.*, Halfhill et al., 2005); however, the effort of producing hybrids and restrictions on their use in uncontained experiments may make simulation experiments simpler and quicker. The weediness or invasiveness potential of transgenic crops themselves can also be tested using experiments that simulate their dispersal to fields of other crops or to non-agricultural habitats (*e.g.*, Raybould et al., 1999).

If experiments falsify a series of hypotheses of no difference in the weediness or invasiveness potential of a transgenic crop, or its hybrids with wild plants, and non-transgenic comparators, the differences could be evaluated using screening tools such as weed risk assessment models (*e.g.*, Bennett and Virtue, 2004; Pheloung et al., 1999; Stone et al., 2008). Weed risk assessment models were developed to assist decision-making about the introduction of new plant species with potential agronomic benefits, such as salt tolerant forage species for southern Australia (Dear and Ewing, 2008; Rogers et al., 2005). The models comprise a series of questions on the likely invasiveness and impact of a species, and use data from the literature, expert judgement and new experimental data if necessary. The models classify potential introduced species into "accept", "further evaluate" or "reject". The models are conservative in that they reject

many species that would probably not become weeds, and in part this is because invasions are rare events (Smith et al., 1999). Statistical methods are being developed to reduce the rate of potentially costly false positives that reject economically valuable species that actually pose low risk (Caley et al., 2006). Treating a transgenic plant as though it were a new species may seem an over-reaction to the addition of one or a few new traits to a very familiar crop, but it would help to put the risks from transgenic crops in context: in effect, the models could be used to test the hypothesis that the weediness potential of the GM crop is no greater than plant species that are well-known not to be weedy or invasive.

Risk characterisation

Transgenic crops are always evaluated for agronomic performance in field trials representing the likely conditions under which the crop will be cultivated. These data allow phenotypic characterisation of the crop (Nickson, 2008) and a test of the hypothesis of no harmful unintended changes in weediness or invasiveness potential as a result of transformation. For crops that are not intended to have enhanced resistance to biotic or abiotic stressors, phenotypic data from agronomic field trials are likely to be sufficient to complete the risk characterisation: many traits associated with weediness or invasiveness potential are likely to show no statistically significant difference from the non-transgenic isoline, and those traits that show statistically significant differences are likely to be in the range of the crop in general. Such results, along with evaluations of the ecotoxicology of the introduced proteins, would strongly corroborate the hypothesis of no harmful unintended changes in weediness or invasiveness potential, and would thereby indicate low risk (Raybould et al., 2010).

In some cases, transgenic crops with enhanced resistance to stressors may require higher tier testing to evaluate phenotypes that are outside the normal range of the crop. Such evaluations would involve experiments to test hypotheses that the intended changes will not increase the abundance of the transgenic crop or sexually compatible wild relatives. Experiments are likely to concentrate on testing hypotheses that the new phenotype will not increase weediness or invasiveness potential; in other words, the studies will test the *effects* of the phenotype. If these hypotheses are falsified, tests of hypotheses about *exposure* to the new phenotype could be carried out; for example, if assessments indicate an increase in weediness or invasiveness potential of a transgenic crop in areas of high salinity, one could test hypotheses that the crop will not disperse to such areas, or that the high salinity areas that it could disperse to do not contain valuable environmental attributes.

The above approaches are quite conservative, in that the baseline comparators are non-transgenic crops of the same species; however, arable crops are rarely serious weeds or invaders (*e.g.*, Warwick and Stewart, 2005). If a transgenic crop consistently shows greater weediness or invasiveness potential than the non-transgenic crop, a less conservative test would be a comparison of the transgenic crop with a wider range of plants using a weed risk assessment model. Such an evaluation would have several useful attributes: it would explicitly include dispersal; abundance would not be the assessment endpoint, as invasiveness (spread) is treated separately from impact (the effect of the plant spreading); and the risk from any increase in weediness or invasiveness potential in the transgenic crop relative to a non-transgenic crop of the same species would be set in context of risks from introducing crop species completely new to an area. It is likely that if any transgenic arable crops were to be characterised as high risk relative to their non-transgenic counterparts, they would be acceptable under weed risk assessments, which themselves have high conservatism as their false positive rates are high (*e.g.*, Caley et al., 2006).

CONCLUSIONS

The bucket and the searchlight are metaphors for opposed theories of scientific knowledge. The bucket theory proposes that knowledge is gained by observing the world without preconceptions; the searchlight theory proposes that preconceptions are unavoidable and that knowledge advances through the improvement of our preconceptions – our hypotheses – by continuous criticism and revision. A hypothesis is a searchlight that seeks to illuminate observations that reveal its flaws, and thereby show how the hypothesis can be improved. In the bucket theory, knowledge emerges from the accumulation of observations that support a hypothesis, whereas in the searchlight theory, knowledge emerges by the elimination of false hypotheses.

The bucket theory offers a plausible explanation for the apparent failure of much environmental risk assessment research to help decision-makers (Craig et al., 2008; Johnson et al., 2007; Raybould, 2007). Researchers may believe that risk assessment research should be objective (Raybould, 2007), and so may make observations about transgenic crops without preconceptions. This is unhelpful for decision making for two reasons: simply collecting observations does not provide predictions about what will happen when a particular transgenic crop is cultivated; and more importantly, risk assessment requires that harmful effects are defined, they cannot be discovered regardless of how many observations are added to the bucket.

The searchlight theory offers a better method for risk assessment. The key to this method is the formulation

and testing of hypotheses that cultivation of the transgenic crop will not lead to harmful effects. Such risk hypotheses are formulated by considering conceptual models that describe plausible pathways by which harm may arise. Risk hypotheses postulate that harm will not arise through a particular pathway, and risk is characterised by testing of those hypotheses. The relevance of a study for risk assessment is therefore judged by its power to refute a risk hypothesis.

Research into the risks posed by the weediness and invasiveness potential of transgenic crops and their hybrids has often forgotten that the purpose of risk assessment is to predict the likelihood and seriousness of harm. Instead it has become side-tracked into measurements of gene frequencies, hybridisation rates and population growth rates, or conceptual discussions about "fitness" (Raybould, 2007; Wilkinson and Tepfer, 2009). A much better approach is to test simple hypotheses that the transgenic crop or its hybrids show no greater weediness or invasiveness potential than non-transgenic plants of the same species, or that the transgenic arable crops have minimal weediness and invasiveness potential compared with other introduced species that have not become weeds. The effectiveness of an environmental risk assessment should not be measured by the size of the bucket of observations on a transgenic crop, but by the power of the risk hypothesis searchlights to clarify the risks that may arise from cultivation of that crop.

REFERENCES

Bennett SJ, Virtue JG (2004) Salinity mitigation versus weed risks – can conflicts of interest in introducing new plants be resolved? *Aust. J. Exp. Agric.* **44**: 1141–1156

Bergelson J (1994) Changes in fecundity do not predict invasiveness: a model study of transgenic plants. *Ecology* **75**: 249–252

Brookes G, Barfoot P (2008) Global impact of biotech crops: socio-economic and environmental effects, 1996–2006. *AgBioForum* **11**: 21–38

Caley P, Lonsdale WM, Pheloung PC (2006) Quantifying uncertainty in predictions of invasiveness, with emphasis on weed risk assessment. *Biol. Invasions* **8**: 1595–1604

Chapman PM, Fairbrother A, Brown D (1998) A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ. Toxicol. Chem.* **17**: 99–108

Craig W, Tepfer M, Degrassi G, Ripandelli D (2008) An overview of general features of risk assessments of genetically modified crops. *Euphytica* **164**: 853–880

Cummings CL, Alexander HM (2002) Population ecology of wild sunflowers: effects of seed density and post-dispersal vertebrate seed predation on numbers of plants per patch and seed production. *Oecologia* **130**: 274–280

Dear BS, Ewing MA (2008) The search for new pasture plants to achieve more sustainable production systems in southern Australia. *Aust. J. Exp. Agric.* **48**: 387–396

Economidis I, Cichocka D, Högel J eds (2010) A decade of EU-funded GMO research (2001–2010). Publications Office of the European Union, Luxembourg

Garcia-Alonso M, Jacobs E, Raybould A, Nickson TE, Sowig P, Willekens H, van der Kouwe P, Layton R, Amijee F, Fuentes AM, Tencalla F (2006) A tiered system for assessing the risk of genetically modified plants to non-target organisms. *Environ. Biosafety Res.* **5**: 57–65

Halfhill MD, Sutherland JP, Moon HS, Poppy GM, Warwick SI, Weissing AK, Rufty TW, Raymer PL, Stewart CN (2005) Growth, productivity, and competitiveness of introgressed weedy *Brassica rapa* hybrids selected for the presence of Bt *cry1Ac* and gfp transgenes. *Mol. Ecol.* **14**: 3177–3189

Hill RA, Sendashonga C (2003) General principles for risk assessment of living modified organisms: lessons from chemical risk assessment. *Environ. Biosafety Res.* **2**: 81–88

Johnson KL, Raybould AF, Hudson MD, Poppy GM (2007) How does scientific risk assessment fit within the wider risk analysis? *Trends Plant Sci.* **12**: 1–5

Keeler KH (1989) Can genetically engineered crops become weeds? *Bio/technology* **7**: 1134–1137

Kessler C, Economidis I eds (2001) EC-Sponsored Research on Safety of Genetically Modified Organisms. Office for Official Publications of the European Communities, Luxembourg

Lawton JH (2007) Ecology, politics and policy. *J. Appl. Ecol.* **44**: 465–474

Lubechenko J (1998) Entering the century of the environment: a new social contract for science. *Science* **279**: 491–497

Marvier M, McCreedy C, Regetz J, Kareiva P (2007) A meta-analysis of effects of Bt cotton and maize on nontarget invertebrates. *Science* **316**: 1475–1477

Maule AJ, Caranta C, Boulton MI (2007) Sources of natural resistance to plant viruses: status and prospects. *Mol. Plant Pathol.* **8**: 223–231

McHughen A, Smyth S (2008) US regulatory system for genetically modified [genetically modified organism (GMO), rDNA or transgenic] crop cultivars. *Plant Biotechnol. J.* **6**: 2–12

National Research Council (2002) Environmental Effects of Transgenic Plants: the Scope and Adequacy of Regulation. National Academy Press, Washington

Nickson TE (2008) Planning environmental risk assessment for genetically modified crops. *Plant Physiol.* **147**: 494–502

Patton DE (1998) Environmental risk assessment: tasks and obligations. *Hum. Ecol. Risk Assess.* **4**: 657–670

Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *J. Environ. Manage.* **57**: 239–251

Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T, Tsomondo T (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agric. Ecosyst. Environ.* **84**: 1–20

Popper KR (1959) The Logic of Scientific Discovery. Hutchinson, London

Popper KR (1979) Objective Knowledge: an Evolutionary Approach. Oxford University Press, Oxford

Raybould A (2005) Assessing the environmental risks of transgenic volunteer weeds. In Gressel J, ed, Crop Ferality and Volunteerism, CRC Press, Boca Raton, pp 389–401

Raybould A (2006) Problem formulation and hypothesis testing for environmental risk assessments of genetically modified crops. *Environ. Biosafety Res.* **5**: 119–125

Raybould A (2007) Ecological versus ecotoxicological methods for assessing the environmental risks of transgenic crops. *Plant Sci.* **173**: 589–602

Raybould A (2010) Reducing uncertainty in regulatory decision-making for transgenic crops: more ecological research or clearer environmental risk assessment? *GM Crops* **1**: 25–31

Raybould AF, Cooper JI (2005) Tiered tests to assess the environmental risk of fitness changes in hybrids between transgenic crops and wild relatives: the example of virus resistant *Brassica napus*. *Environ. Biosafety Res.* **4**: 125–140

Raybould AF, Wilkinson MJ (2005) Assessing the environmental risks of gene flow from genetically modified crops to wild relatives. In Poppy GM, Wilkinson MJ, eds, Gene Flow from GM Plants, Blackwell Publishing, Oxford, pp 169–185

Raybould AF, Moyes CL, Maskell LC, Mogg RJ, Warman EA, Warlaw JC, Elmes GW, Edwards M-L, Cooper JI, Clarke RT, Gray AJ (1999) Predicting the ecological impacts of transgenes for insect and virus resistance in natural and feral populations of *Brassica* species. In Ammann K, Jacot Y, Simonsen V, Kjellsson, eds, Methods of Risk Assessment of Transgenic Plants, III. Ecological risks and prospects of transgenic plants, Bïrkhäuser Verlag, Basel, pp 3–15

Raybould A, Stacey D, Vlachos D, Graser G, Li X, Joseph R (2007) Non-target organism risk assessment of MIR604 maize expressing mCry3A for control of corn rootworm. *J. Appl. Entomol.* **131**: 391–399

Raybould A, Tuttle A, Shore S, Stone T (2010) Environmental risk assessments for transgenic crops with enzyme-based output traits. *Transgenic Res.* **19**: 595–609

Raybould A, Caron-Lormier G, Bohan DA (2011) The derivation and interpretation of hazard quotients to assess the ecological risks from the cultivation of insect-resistant transgenic crops. *J. Agric. Food Chem.*, DOI: 10.1021/jf1042079

Rogers ME, Craig AD, Munns RE, Colmer TD, Nichols PGH, Malcolm CV, Barrett-Lennard EG, Brown AJ, Semple WS, Evans PM, Cowley K, Hughes SJ, Snowball R, Bennett SJ, Sweeney GC, Dear BS, Ewing MA (2005) The potential for developing fodder plants for the salt-affected areas of southern and eastern Australia: an overview. *Aust. J. Exp. Agric.* **45**: 301–329

Romeis J, Bartsch D, Bigler F, Candolfi MP, Gielkens MMC, Hartley SE, Hellmich RL, Huesing JE, Jepson PC, Layton R, Quemada H, Raybould A, Rose RI, Schiemann J, Sears MK, Shelton AM, Sweet J, Vaituzis Z, Wolt JD (2008) Assessment of risk of insect-resistant transgenic crops to non-target arthropods. *Nature Biotech.* **26**: 203–208

Sanvido O, Romeis J, Bigler F (2007) Ecological impacts of genetically engineered crops: ten years of field research and commercial cultivation. *Adv. Biochem. Engin. / Biotechnol.* **107**: 235–278

Smith CS, Lonsdale WM, Fortune J (1999) When to ignore advice: invasion predictions and decision theory. *Biol. Invasions* **1**: 89–96

Squire GR, Brooks DR, Bohan DA, Champion GT, Daniels RE, Haughton AJ, Hawes C, Heard MS, Hill MO, May MJ, Osborne JL, Perry JN, Roy DB, Woiwood IP, Firbank LG (2003) On the rationale and interpretation of the Farm Scale Evaluations of genetically modified herbicide-tolerant crops. *Phil. Trans. R. Soc. Lond. B* **358**: 1779–1799

Stone LM, Byrne M, Virtue JG (2008) An environmental weed risk assessment model for Australian forage improvement programs. *Aust. J. Exp. Agric.* **48**: 568–574

Suter GW (1990) Endpoints for regional ecological risk assessments. *Environ. Manage.* **14**: 9–23

Suter GW (1996) Abuse of hypothesis testing statistics in ecological risk assessment. *Hum. Ecol. Risk Assess.* **2**: 331–347

Sutherland J, Poppy G (2005) Risk assessment of *Bacillus thuringiensis* in wild *Brassica rapa*: a field simulation of introgressed transgenes. *J. Plant Interact.* **1**: 31–38

Touart LW, Maciorowski AF (1997) Information needs for pesticide registration in the United States. *Ecol. Appl.* **7**: 1086–1093

Warwick SI, Stewart CN (2005) Crops come from wild plants – how domestication, transgenes and linkage together shape ferality. In Gressel J, ed, Crop Ferality and Volunteerism, CRC Press, Boca Raton, pp 9–30

Warwick SI, Beckie HJ, Hall LM (2009) Gene flow, invasiveness, and ecological impact of genetically modified crops. *Ann. N.Y. Acad. Sci.* **1168**: 72–99

Wilkinson M, Tepfer M (2009) Fitness and beyond: preparing for the arrival of GM crops with ecologically important novel characters. *Environ. Biosafety Res.* **8**: 1–14

Wolt J, Keese P, Raybould A, Fitzpatrick JW, Burachik M, Gray A, Olin SS, Schiemann J, Sears M, Wu F (2010) Problem formulation in the environmental risk assessment for genetically modified plants. *Transgenic Res.* **19**: 425–436