Final Draft. To appear in: Breyer, T. (2013). *Epistemological Foundations of Evolutionary Psychology*. New York: Springer.

When misrepresentation is successful

Michael Zehetleitner
Ludwig-Maximilians-Universität München, Munich, Germany

Felix D. Schönbrodt
Ludwig-Maximilians- Universität München, Germany

Author Note

Michael Zehetleitner, Felix D. Schönbrodt: Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany.

Correspondence concerning this article should be addressed to Michael Zehetleitner, Leopoldstr. 13, 80802 München, Germany. Email: mzehetleitner@psy.lmu.de.

Phone: +49 89 2180 5209. Fax: +49 89 2180 5211.

Abstract

In the present chapter, we investigate the notion that an action is successful if and only if (iff) it is caused by a true representation. We demonstrate that there indeed exist representations which - even though being false - can systematically lead to successful actions, if specific conditions hold, especially, if there is stochastic noise in the generation of representations and the cost of errors is asymmetrically distributed and the success-relevant feature can only be indirectly assessed via indicator features. Finally, we discuss this observation in relation to illusionary perception and evolutionary epistemology.

Keywords: Semantics, success semantics, evolutionary epistemology, teleosemantics, decision theory

When misrepresentation is successful

How can a representation lead to a successful action? It is widely taken for granted that a representation has to be true in order to be successful (e.g., Shea, 2007; Millikan, 1989; Whyte, 1990; Ramsey & Moore, 1927; Blackburn, 2005). Consider for instance, how could a person successfully sit down on or avoid a chair, unless she has a true visual representation of the chair's shape and position? Or, to make use of an example of Ramsey in Ramsey and Moore (1927): How can the belief of a chicken, that a certain caterpillar is toxic be useful, unless the caterpillars are actually toxic? It even has been proposed, that "(t)ruth just *is* the property of a belief that suffices for your getting what you want when you act upon it." (Whyte, 1990, p. 149). Although this notion of truth being a prerequisite for success has a high face validity, we shall argue that it is wrong to suppose that all successful actions require true representations. Success does not require truth in all and every cases. Under certain circumstances, false representations can systematically cause successful actions. The goal of the chapter is to demonstrate what these defined circumstances are.

Success Semantics

Throughout this chapter, we assume a naturalistic theory of semantics, where representations, their content, their truth or falsehood are defined without recurrence to terms which themselves are already intentional. Our argumentation is invariant to what exact version of naturalized semantics is assumed (e.g., Fodor, 1990; Millikan, 1989; Papineau, 1984, 2003; Dretske, 1981). Now, what does it mean for a representation to be true vs. false or successful vs. unsuccessful?

Representation

We understand that observer *O* has representation *r*, if and only if (iff) *r* is a (physical or biological) state or signal within *O*, which has the property of being about something else, that is to have content. To have content can be considered as a mapping from a set of representations *R*[1]

[1] Capital letters here denote variables. A variable is a set of possible values together with a measurement operation which allows determination of what value currently is the case. For instance, the size of an apple can be considered as a variable S. The possible values of S are within the interval between zero and infinity. The measurement is a ruler, which also determines the quality of the variable, namely [cm] or [m].

(from which r is an element of) to a specific set of states in the world external to the observer, the target domain T.[2]

$$cont: R \rightarrow T. \quad (1)$$

Together with the content mapping, when an observer *O* has representation $r \in R$ and $cont(r) = t^*$ it is possible to state: observer *O* has a representation *r* the content of which is t^*, or O believes that T is t^*.[3] Using this notation for content, it is possible to dissociate two aspects of 'aboutness' (or 'intentionality'): first, a set of representations *R* is about a specific set of world states T (the domains of the content mapping *cont*) and not about a different set of states, say *T'*. For instance, a specific representation *R* is about the velocity of an object and not about its size. This first aspect of 'aboutness' specifies which measurement unit the representation's content has (e.g., [m/sec] or [m]). Second, a specific representation $r \in R$ has the specific content $cont(r) = t^* \in T$. To use a neurophysiological example, neuronal activity in orientation columns in primary visual cortex (of e.g., cats, or primates, including humans) would be a set of representations *R*, which is about the angular orientation (e.g., vertical or horizontal) of line segments and not about their colour (e.g., Hubel & Wiesel, 1974). A specific pattern of neuronal activity then represents a specific angular orientation (e.g., horizontal) and not any other orientation (e.g., vertical).

Truth

In order to talk about the truth or falsehood of a representation, it is required to determine one further component: the actual state of the world, $\underline{t} \in T$. Remember, when the content of a certain cell assembly in a cat V1 is "horizontal" (i.e., 0°) it is possible to state that the cat believes that the bar is horizontal. This belief is true, in case the actual state of the world (that is the line segment's orientation) indeed is "horizontal" ($\underline{t} = t^*$), and false, in case it is not ($\underline{t} \neq t*$):

$$\text{Representation } r \in R \text{ is true} := cont(r) = \underline{t}. \qquad (2)$$

Success

[2] Here, we consider only first order representations, the content of which is situated in the observers' external world. In principle, representations can also be meta-representations, i.e. about other representations.

[3] In the present chapter, we consider both statements to be synonymous, which need not necessarily be the case.

Now, when is a representation successful? A representation is considered successful if it is useful for actions (Ramsey, 1927), it allows desires to be fulfilled (Blackburn, 2005; Whyte, 1990), or if it is causally effective in increasing the organism's fitness, i.e., the expected value of the number of reproducible offspring (e.g., Millikan, 1989). A frequent assumption is, that representations are successful if and only if they are true: for any representation $r \in R$, and the actual state of affairs $\underline{t} \in T$, an operation to determine the success of action *a*, *successful*, and the content mapping *cont* between *R* and *T*:

$$(successful(a) \wedge r \rightarrow a) \Leftrightarrow cont(r) = \underline{t}, \qquad (3)$$

that is the content of the representation is equal to the actual state of the world (2).[4]

Until here, truth and success have been introduced as binary. However, the terminology until now can also include gradual cases: for instance, a deer wants to jump over a 1.2 meters wide ditch when trying to evade a predator. If its internal representation of the ditches width is true, it can plan its jump and successfully reach the other bank. If its internal representation however is false, it can be so in many ways. The width-representation's content can be for instance be 1.199 meters, which is false. Still, the evading jump could be successful. If the representation's content is for instance 1.1 meters, the deer could still evade the predator, but twist its ankle. If it's content is 0.5 meters, its desire to evade the predator will most probably turn out not to be fulfilled. This example illustrates that for states of the world *T* for which an ordinal scale of measurement exists, the deviation from truth can be rank ordered, i.e. for two false representation one can be "more false" than the other. That is, even though logical truth is still binary, the falsehood of representations can be gradually qualified and differentiated. Also success can be used as a binary term. Therefore, it is necessary, to introduce a further term, which quantifies the outcome of each action combined with each state of the world *T*. Applied to the deer example, a jump of any length *a*, under the condition of any width of ditch, *t*, has a certain consequence, hence termed utility.[5] Thus, for a given length of jump *a*, the result differs in utility depending on the actual width of the ditch *t*. Utility should be maximal for that width of ditch, which corresponds to the length of jump, i.e., for *a*=*t*. An action a is called successful, if its result has the highest utility, given the current state of affairs, t, compared to all other possible actions.

[4] Although it might be debated whether the truth of a representation "guarantees" success (see Nanay, 2012; Blackburn, 2005), there is general consensus that a false representation is incapable of generating a successful action other than accidentally.

[5] Utility is used in a similar sense in evolutionary biology (e.g., Reeve & Sherman, 1993; Fox & Westneat, 2010). In decision science it is frequently termed pay-off, (cost) value, or reward (e.g. Green & Swets, 1966; Gold & Shadlen, 2007). It can be quantified in cardinal or ordinal scale.

A requirement is that utility can be assigned an at least ordinal scale, where the outcome of an action can differ in its results. For instance, biological fitness meets this requirement. The relationship between truth and success (3) in this case would state that a representation is true *iff* it causes the most utile action.

Consequently, independent of how utility is defined in detail, it at least comprises a measure which can be assigned to each action/world combination, because the utility of an action is always relative to in what state the world currently is. Applied to the deer example, for each width of ditch (states of the target domain *T*) and each jump length (set of actions *A*) a (for instance real) measure can be assigned quantifying the action's utility given the current state of the target domain:[6]

$$\mathit{\Omega}\colon A \times T \to \mathbb{R};\ (a,t) \mapsto \omega(a,t). \qquad (4)$$

For the deer example the utility surface is depicted in Figure 1. For each state of the world, T, the width of ditch in the deer example, utility is maximal for a jump of the corresponding length (i.e., where the width representation is true), which is in line with success semantics (3) the successfulness of an action, *successful*, in Eq. (3) is defined using the construct of utility. An action $a' \in A$ is successful for a given $t \in T$ if a' has the highest utility value compared to all other $a \in A$ given *t*:

$$successful(a')\colon a' = argmax_a \omega(t,a). \qquad (5)$$

[6] Assuming that a representation invariably causes the same action, in the present case, a representation of x meters always leads to a jump of x meters.

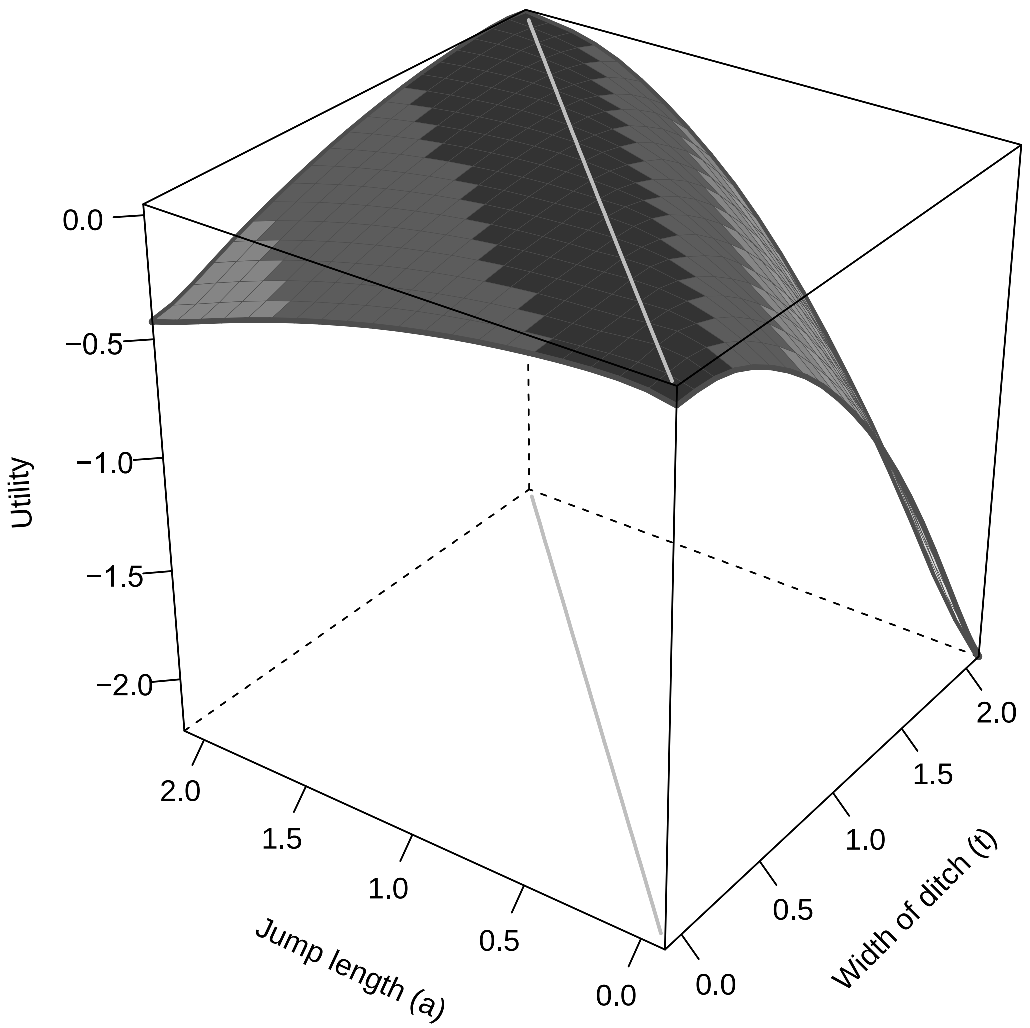


Figure 1. This (hyper-)surface denotes the utility relationship between width of ditch, t (x-axis) and length of jump a(y-axis) on utility (vertical z-axis). One can observe that for each width of ditch, there is one length of jump, for which utility is maximal, represented by the white line on top of the surface and its projection onto the T x A plane. This maximal utility (i.e., success) is achieved for correspondence between width of ditch and jump, respectively, i.e. a=t. Jumping too short (a<t) leads to less utility than jumping too far (a>t), the latter only causing waste of energy, the former potential injuries or capture. The surface is not restricted to continuous variables of T and action A. For e.g. dichotomic variables, the surface would consist of four points where again for each action the largest utility value is present at the true representation (see for instance Table 1 below).

Now, if there were a one-to-one mapping between the target domain and representations, all representations would always lead to maximal utile, i.e., successful actions, because every ditch width *t* would invariably lead to the true representation *r*. That is, there would be no chance of misrepresentation. In such a situation, in information theoretic terminology, the mutual information between the target domain *T* and the set of representations *R* would be equal to *T*'s entropy, that is, there would be no loss of information. Generally, information between *T* and *R* is lost, if *R* is not solely causally affected by *T* but also by any other source, *Z* (see Figure 1).

Optimal bias: no challenge to success semantics

Let's apply this abstract conception to an example provided by Godfrey-Smith (1991), which has been first discussed by Millikan (1989; see also Usher, 2001) and which has been considered as a potential counterexample for success semantics. A hunting tiger sees a movement of the grass. Either, the grass is moved by prey or by the wind and the carnivore has to make the decision whether to jump and attack or not to. That is, the target domain set *T* consists of the elements {prey, non-prey}, the set of actions *A* of the elements {jump, ignore}.

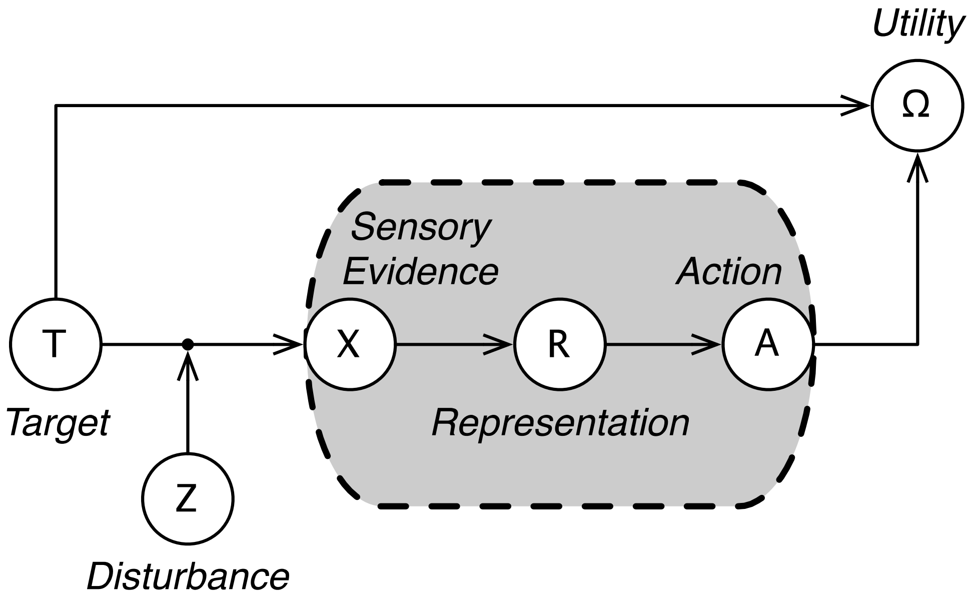


Figure 2. Variables are denoted by circles, arrows denote causal connections. The dashed line denotes the boundaries of the organism O, which has representation R. R is about the state of the world T, external to O. T affects O's sense organs producing sensory evidence X. R causes one of a possible repertoire of actions A. Ω denotes the utility of each action A combined with each state of the world T.

Let *X* denote a signal in the tiger's visual system, the strength of which is proportional to the vehemence with which the grass is disturbed. The outcome of the tiger's decision about presence of prey is a representation *R*, the content of which is *T*, that is either prey or non-prey. As a simplification, following Godfrey-Smith, let's assume that there is an invariable link between *R* and the tiger's action *A*: it always jumps when it believes that prey is present and ignores non-prey. How does the utility surface look like? First, the surface can be represented as a matrix, because both *T* and *A* are dichotome. Jumping when prey is present ultimately leads to feeding which here is assigned +20 in arbitrary units of utility. Jumping when no prey is present leads to a waste of energy and is assigned a utility of -5. The utility of ignoring is independent of whether prey is indeed present or absent and set to -0.1[7]. This reflects the fact that due to metabolic loss of energy continuous ignoring will lead to starvation. In general, the utility matrix for combinations of the target domain *T* and actions *A* is:

	Action	
Target	Jump	Ignore
Prey	+20 (hit)	-0.1 (miss)
Non-prey	-5 (false alarm)	-0.1 (CR)

Table 1. Utility values ω for each combination of t and a for the tiger example illustrated above.

Now, not only prey can disturb the grass, but also the wind. That means the information flow between presence of prey *T* and the sensory state *X* is reduced by the disturbance (denoted Z in Figure 2). The disturbance can have a source external to the organism (e.g., wind moving the grass) or internal to it (i.e., neuronal noise). In case no information is lost, there are no errors. In case of information loss, the animal has to balance which type of mistakes to make: Should it be unbiased and equally often jump at non-prey as it ignores prey? Such a decision scheme would intuitively seem disadvantageous, given that jumping at non-prey is more costly[8] than ignoring prey (see Table 1). Indeed, decision theoretic calculations confirm that the tiger acts optimal if it is biased towards jumping rather than ignoring, because the cost of a false decision is greater in case prey indeed is present compared to when prey is absent (e.g., Godfrey-Smith, 1991). Thus, an optimistic tiger throughout life gains more calories than its sceptic fellow tiger.

The opposite is true for beavers (using an example of Millikan, 1989): optimistic beavers which, when the grass is disturbed, assume it to be the wind rather than a predator have a shorter life expectancy than skittish beavers which raise alarm and hide at the slightest disturbance of

[7] The exact numerical values are arbitrary.

[8] Costs can be defined in two ways. First, negative entries in the utility matrix can be considered costs. The definition we are using in the subsequent chapter is the following: The incorrect action has less utility than the correct action. We understand as cost the amount of how much less utility an incorrect action has, compared to the correct action. That is, the ‘costly’ action can have a positive entry in the utility matrix, but we would still speak of cost, as utility would be positive but smaller than for the correct action.

grass. Here, the cost of a false decision in case a predator is present is greater than the cost in case no predator is present.

Decisions can be (optimally) shifted not only depending on the cost functions, but also on the base probability of events *T*. Consider a bear that is foraging at a river full of salmon, optimally has a very liberal criterion and claws the water at the merest indication of a flicker in the water. When fish are scarce, it is optimal to require more visual evidence for making a fishing attempt (for optimal foraging behaviour see, e.g., Stephens & Krebs, 1987).

The conclusion to be drawn from these examples is twofold. First, there can be calculated decision criteria, which are optimal in the sense that they maximize the expected value of utility, taking into account the cost of different types of errors and the a-priori frequencies within the target domain.[9] Second, this optimal criterion can be biased favouring one type of mistake (false alarm or miss) over the other, in cases, where both errors differ in cost, or where the a-priori base rate of events is not uniform.

Would such optimally biased decision criteria challenge success semantics? One could argue that a bear foraging at a river rich in fish should be credulous rather than sceptic. It's overall utility is maximized if it frequently falsely believes fish to be present rather than impartially or sceptically evaluating visual evidence. Couldn't one now argue that the falsity of the bear's beliefs leads to maximal utility? Although this argument to our knowledge has never been proposed, it has been discussed and refuted (Millikan, 1989; Godfrey-Smith, 1991; Usher, 2001). The refutation argument is as follows: A high proportion of false beliefs can lead to maximal utility in case a lot of situations are aggregated (i.e., *in average*). However, in each *specific instance*, the bear can feed iff its belief about the presence of fish was true. Also, each false alarm in every *specific instance* leads to a loss of energy. The fact that on average false alarms are rare, because the river is packed with fish, thus cannot challenge the close coupling between truth and success of a single representation. This coupling is reflected in the utility matrix, where for each action the utility is maximal iff the representation is true, and the utility matrix itself remains unaltered even if the a-priori base frequencies of events change. Consequently, even though frequent misrepresentations are optimal, they are no challenge to success semantics (3).

[9] Please refer to Godfrey-Smith (1991) for the analytic derivation of the optimal decision criterion. An alternative (and equivalent) approach has been described by Bischof (1998).

False representations which lead to success

Here, we demonstrate that it is indeed possible that a false representation systematically causes successful actions - not only on average, but in specific instances (building upon an example first presented by Bischof, 2009). Above, we have summarized that frequent false representations are optimal (i.e., maximizing average utility) under specific conditions: first, there has to be information loss between the target domain and the biological signals, based on which representations are formed. Second, there has to be present at least one type of asymmetry: either, the a-priori frequency of events in the target domain is not uniform, or the utility matrix Ωhas to be asymmetric. We now argue that in cases, where actions are based on indicator representations, success semantics can be violated.

Indicator representations

In order to describe what we mean with indicator representations, let's sketch out a further example of monkeys in presence of toxic and harmless snakes (cf. Bischof, 2009). First, without considering indicator representations, the target domain of the representation is the snake's toxicity, T_t ={toxic, non-toxic}. The monkey's representation R_t of the fact also is binary and can cause the actions A={eat, run} (see Figure 3).

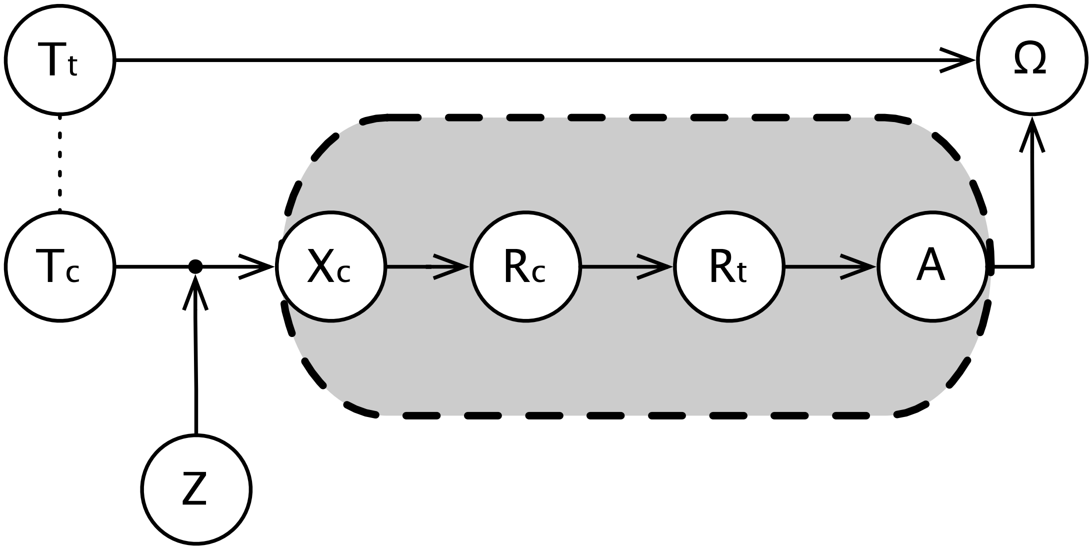


Figure 3.The notations is equivalent to Figure 2. New variables are T_c, and R_c, which denote a state of the world external to the organism and its representation. The dotted line between T_c and T_t indicates at least a correlational relation, possibly a causal connection.

The utility matrix is represented in Table 2. Trying to eat a toxic snake is not assumed to be deadly, but to have severe fitness consequences. Eating a snake has moderate gain. Here, success semantics (3) is satisfied: for each action utility is maximal iff the mediating representation is true.

	Action	
Toxicity	Eat	Run
toxic	-1	0
Non-toxic	0.5	0

Table 2. Utility matrix Ω omega for the monkey example illustrated above.

Now, the monkey does not have a poison detector, but snakes come in different colours and nature has it that toxic snakes are coloured bluish, whereas harmless snakes are of a greenish colour. That is, the snakes' colour can be used as an indicator for their toxicity. The snakes' colour T_c is assumed to vary continuously between saturated blue and saturated green, including a completely desaturated (i.e., greyish) colour. Therefore, T_c varies between -∞ and ∞, where 0 stands for grey, negative values for a blue hue, positive values for a green hue, and its absolute value for the colour's saturation. As the monkey has a colour detector, the activity of which is denoted X_c (see Figure 3), it still is able to distinguish toxic from harmless snakes and act accordingly. However, again, there is information loss between the actual colour T_t and sensory evidence X_c. Specifically, the sensory evidence for each colour t_c is assumed to be normally distributed around x_c. As a consequence, there is uncertainty, given a specific value of sensory evidence, what the snake's colour actually was (see also Figure 4, left panel).

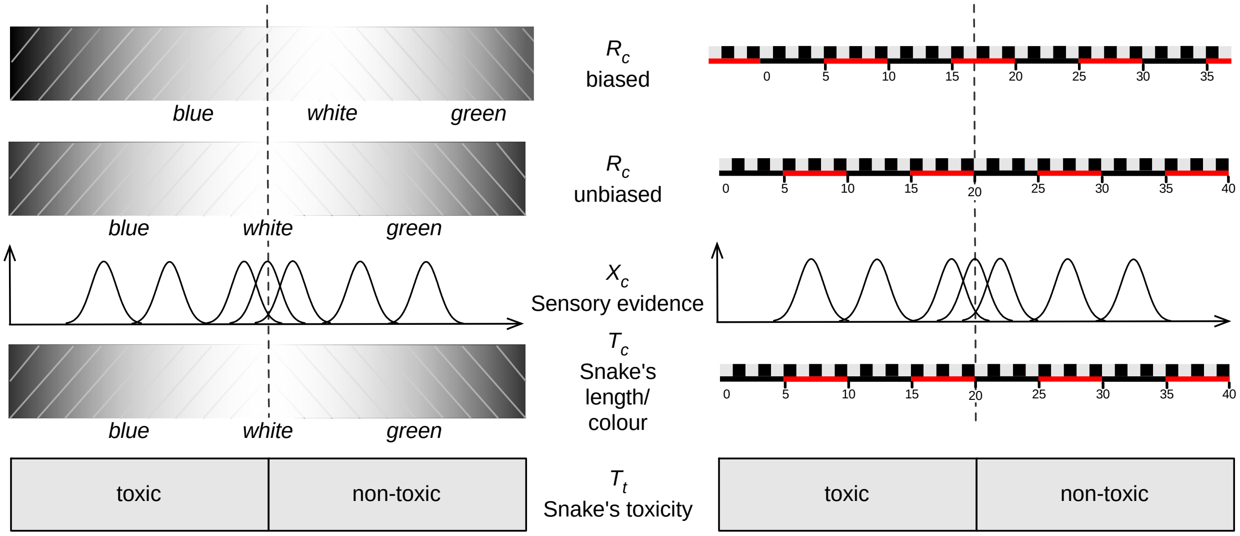


Figure 4.The left panel presents toxicity T_t, colour T_c, and sensory evidence X_c according to the thought experiment. T_t and T_c are perfectly correlated with snakes of a completely desaturated colour being non-toxic. In the figure, the snake's colour's hue is represented by texture and saturation by luminance. Toxic snakes have a blue hue (right tilted texture) and non-toxic snakes have a green hue (left tilted texture). White (and white) colour in the figure denotes completely desaturated (saturated) colour of the snake. The top of the figure displays two possibilities how a colour representation R_c can be formed based on X_c: unbiased, where the colour grey leads to representations with content blue and green with equal probability, or blue-biased, where grey mostly leads to representations with content blue. The right panel presents an analogous example, where snakes differ in size and snakes smaller than 20 cm are toxic. R_c are again presented for the unbiased and the biased case.

But how does the monkey decide upon its action given its visual input? First, it forms an indicator representation of colour, $R_{c,}$ about T_c, based on sensory evidence X_c and finally, depending on the colour representation, a representation of the toxicity R_t is triggered (with blue snakes always being treated as toxic, that is leading to a 'run' action). A snake is assumed to be toxic, if the colour representation indicates a blue hue, i.e. for all $r_c<0$ representation r_t = toxic is formed. For all $r_c \geq 0$, the representation r_t = non-toxic is formed. The whole setup is depicted in Figure 3.

This setup differs from the thought experiments presented above in one important aspect: it is no longer assumed that there is a direct causal connection from the fitness relevant feature (the toxicity of the snake) to the sensory apparatus of the monkey. Rather, a different feature, in this case colour, is picked up by the observer and used as an indication for the success relevant feature, which is possible if there is a correlation between the indicator feature T_c and the success relevant feature T_t. The subsequent discussion does not depend on there being no causal connection between T_t and T_c. It could well be that the poison within the snake causes the production of a molecule within the snake's skin, making it appear blue. The crucial difference is that there is no direct causal route from the success relevant feature T_p to sensory evidence, but only an indirect one via T_c (and thus X_c).

Plausibility check

Before discussing the impact of indicator representations on success semantics, let's first check their plausibility. In fact, there are natural examples in abundance, where indicator representations are used by organisms, because frequently they do not possess sense organ which are directly causally affected by those features which are relevant for the success/utility of actions such as fertility of a mate, nutritiousness of food, danger of a predator. Consider for instance the tiger in the example of Godfrey-Smith (1991): it used disturbances of the grass as an indicator for presence of prey. Consider alternatively the famous example of Dretske (1986): there exist prokaryotes which have a cell organelle responsive to magnetic north. Those anaerobic prokaryotes with a habitat in the northern hemisphere align themselves to drift north, in the southern hemisphere to drift south (Blakemore, 1975; for a recent review see Bazylinksi & Frankel, 2004). In both habitats the organisms drift into deep waters which contain less oxygen. Here, magnetic north is used as an indicator for oxygen, because the utility of their alignment action depends on the oxygen level of the waters they are drifting into, not on their magnetic properties. Let's briefly touch a couple of further examples. Frogs and toads are known to snap at small dark moving objects or longish things moving along their long axis, respectively (Lettvin et al., 1968; Ewert, 1974; Borchers, Burghagen, & Ewert, 1978), indicating presence of prey. Sickle-back males show behaviour of territory defense when confronted with an oval, medium sized dummy with a red lower half (Tinbergen, 1951) and sickle-back females show courtship behaviour for dummies of correct hue, contrast, and configuration (Baube, Worland, & Fowler, 1995), that is they use visual configurations as indicators for presence of rivals or mates. Finally, vervet monkey sentinels can issue alarm calls based on which other vervets in the vicinity perform an evading action fitting to the type of attacker indicated by the alarm call. In this case, the reacting vervets have potential sensors for detecting predators, but they use the alarm call as an indication of predators' presence, even though these are currently out of their sensory range (e.g., Seyfarth, Cheney, & Marler, 1980).

Example	Action	T_t: Success relevance	T_c: Indicator
monkey	locomotion	toxicity	colour
tiger	attack	nutritional value	visual configuration
beaver	locomotion	danger of bodily harm	visual/auditory configuration
magnetotactic bacteria	locomotion	oxygen	magnetic north
frog	feed	nutritional value	visual configuration
toad	feed	nutritional value	visual configuration
sickle-back male	defend	competitor for resources	visual configuration
sickle-back female	court	fertility of mate	visual configuration
vervet	locomotion	danger of bodily harm	alarm call

Table 3. Summary of biological examples in which successful actions are triggered by indicator representations.

Each of these examples (summarized in Table 3) can be mapped upon the architecture as presented in Figure 3: there is a success-relevant variable T_t which there is no direct sensor, which T_t would directly causally affect.[10]

However, there is a sensor X_c for something else, an indicator feature T_c, which is frequently correlated with T_t . In each of these cases, the class of objects triggering an action ("real things plus dummies") is greater than the class for which the reaction leads to success ("only real things"). Snapping at small dark moving objects does not feed the frog - at flies does. Also, in all of these cases, in the environment of evolutionary adaptiveness, frequently presence of the indicator coincides with presence of the success-relevant object. Importantly, the success of actions triggered by R_c does depend solely on T_t, not on T_c. To refer to the main example of this chapter, the success of running away only depends on the snake's toxicity - not on its colour.

In sum, the present setup exemplifies a rather large class of cases, where a success-relevant variable is picked up by an organism only indirectly, using an indicator variable and thus the consequences for success semantics, as developed below, are relevant for a rather large class of cases as well.

[10] In the following, we keep the indices c for indicator and t for success-relevant representations maintaining continuity with the main example without implying restriction to that specific example.

Optimal mutations

As outlined above, in the presence of information loss between the target domain and sensory evidence, and if costs of misrepresentations are unequal or the a priori probabilities within the target domain are not uniform, average utility can be optimally increased, when there are frequent misrepresentations. In the setup discussed until now (e.g., Godfrey-Smith, 1991, see Figure 2) misrepresentations could be introduced at the transformation of sensory into a representation. In the presence of indicator features/representations, there are two possible loci, where misrepresentations can be introduced: first, as in the simpler setup, at the transition to R_t, or second at the transition from X_c to R_c. The relation to success semantics of these two possibilities will be discussed in turn.

R_t Shift

As in the absence of an indicator representations, introduction of misrepresentations between the representation R_c of the indicator, T_c, and the representation R_t of the success-relevant domain T_t, can increase average utility: treating even slightly greenish snakes as toxic may be unsuccessful in each specific instance, but on average, utility can be maximised, because less errors of the costly type (trying to eat a toxic snake) are less frequent than errors of the harmless type (running away from a non-toxic snake). In this case, success semantics still holds, because each specific action is successful iff R_t is true.

R_c (Indicator) Shift

Crucially, consider that there are genetic variants of the monkey, shifting its colour representation R_c. These variants, when encountering a grey snake, would believe it to be of a certain shade of blue, in case R_c is shifted towards a bluish bias, or would believe it to be of a certain shade of green, in case R_c is shifted towards a greenish bias. The saturation of R_c when encountering a grey snake would correspond to the colour-bias ranging from some negative to some positive constant. What of these variants would have the greatest fitness, the highest number of surviving offspring? Let's first consider a slight greenish bias. Then, slightly bluish snakes ($T_c < 0$) would lead to greenish colour representations ($R_c > 0$). Such a monkey would frequently try to eat toxic snakes - especially more frequently than its relative which has an unbiased colour representation. Let's second consider slight bluish bias. Such a monkey would frequently believe a slightly greenish (and thus harmless) snake to be blue, and thus try to evade it. However, it is conceivable, that due to that slight misrepresentation of colour, it makes more mistakes trying to eat slightly bluish snakes, as it's colour representation R_c amplifies the actual blue colour of the snake (T_c). Even though it thus misses more slightly greenish snakes, overall, it should have a higher accumulated utility than its unbiased relative, because mistaking a green colour for blue has more harmless consequences, than mistaking a blue snake for green. Further mutations, which have very strong bias of blue, would try to evade nearly snake, except the deep green ones. Although such a monkey would never get bitten by a toxic snake, it also would fail to feed on the harmless snakes. Consequently, there seems to be some slight bias of bluish tint for R_c, which would have the highest fitness value: higher than unbiased mutations and higher than mutations with an even stronger bias of blue. Before quantitatively demonstrating that this argumentative, qualitative reasoning is indeed correct with a computational, evolutionary simulation, let's explore the consequences for success semantics.

Consequences for success semantics

For the optimal misrepresentation of colour, i.e. grey having a slightly bluish tint, so far, success semantics still holds, because misrepresentations of R_c in these cases leads to misrepresentations of R_t and consequently to unsuccessful actions, in each specific instance. Crucially, however, as both the indicator feature T_c, and its representation R_c are continuous variables, there are cases, where blue snakes are misrepresented as bluer than they actually are, and green snakes as less green than they are. In both of these cases, toxic snakes lead to a successful evasion action and non-toxic snakes are eaten. That is, even though the colour representation is false, the resulting action is successful, in a systematic way - violating the core assumption of success semantics (3) that in order to systematically cause a successful action, a representation has to be true.

Let's be even more concrete. The misrepresentation of colour can be quantified with a value of b, such that the content of the colour representation is that of the actual colour, shifted by b:

$$cont(r_c)=t_c+b,$$

for a representation $r_c \in R_c$. That means, a completely desaturated colour, t_c=0, leads to a representation $r_c=b$, which is true in case b=0, and which corresponds to saturation corresponding to the absolute value $|b|$ of b of green in case b>0 and of blue in case b<0.

Consequently, colours in the range of $-b<t_c<b$ lead to unsuccessful misrepresentations R_c, because green snakes are mistaken as blue, and vice versa, and consequently toxic snakes are treated as non-toxic, and vice versa. However, for snakes of colour $|t_c|>b$, even though the colour representation is wrong (indeed wrong by the amount b), green snakes are treated as harmless and blue snakes as toxic. The exact value of b of course depends on the probability distribution of T_t, the asymmetry of the utility matrix, and the amount of information loss between T_c and X_c. The following section presents an evolutionary simulation of the monkey/snake example in order to validate the qualitative argumentation presented above.

Simulation

Description of the situation

Following simulation will be used to illustrate the impact of information loss (e.g., sensory noise), asymmetry of the utility matrix, and misrepresentation on the utility of an action. The scenario of monkeys and snakes given above will be used as illustration. Recall that the monkey can encounter *non-toxic snakes*, which can be eaten and increase the monkey's reproductive success. The *toxic snakes* attack monkeys that try to eat them with a poison that makes them infertile for some time. The skin colour can be used as a cue for the type of a particular snake: the skin of healthy snakes has a slight green tint, and the skin of the toxic snakes has a slight blue tint (see Figure 4).

Some monkeys are born with a minor genetic defect that causes them to misperceive the colour spectrum in a way that everything gets a blue tint. In this case, white will be perceived as blue, which in the case of snakes is a conservative perception (the monkey rather runs away from a white healthy snake than getting bit). The simulation aims at demonstrating the impact of sensory noise and the amount of misrepresentation on the fitness of the monkeys.

Method and Results

As an indicator for the monkey's fitness the number of expected descendants is computed ("lifetime reproductive success"; Abrams, 2012). Following numbers were used for the simulation:

- Maximum life duration of a monkey: 5 years

- Average number of descendants in the monkey population: 1 per year
- Encounters with a snake (of unknown type): once a year
- Probability of encountering a toxic snake: 50%
- Trying to eat a toxic snake leads to -1 descendants in that year (resp., in the "strong toxic" case, to -2 descendants)
- Eating a healthy snake increases the number of descendants in this year by 0.5

The simulation had two experimental factors. On the one hand, four subpopulations of monkeys with different sensory fidelity were simulated by adding Gaussian noise with M=0 and SD = 0, 0.1, 0.2, 0.3 to the true colour signal[11]. On the other hand, the level of misrepresentation was varied by inducing a sensory shift in R_c ranging from b = -0.6 to b = 0.2. Negative levels correspond to the induction of a blue tint, leading to the misperception that a whitish healthy snake (with a colour value of 0.05, for example) is perceived as a toxic snake.

For each experimental condition, the life courses of 10'000 monkeys were simulated and the average number of descendants computed as a measure of fitness. The results are displayed in Figure 5. The x-axis shows the amount of colour misrepresentation, the y-axis the fitness, and different line types show different levels of noise. The left panel shows a situation with high asymmetry of utility, and the right panel low asymmetry. The asterisk marks the optimal sensory shift for each level of sensory noise.

One can clearly see that for noise levels > 0 the optimal fitness is achieved with a misrepresented colour. Increasing noise levels lead to increasing shifts in sensory perception. Not surprisingly, the overall maximum of reproductive success is achieved with the most accurate perception (lowest noise). Hence, selection pressure should favour better sensors, and better sensors should go along with a more accurate representation.

[11] The general conclusions do not dependent on the specific numbers used in the simulation.

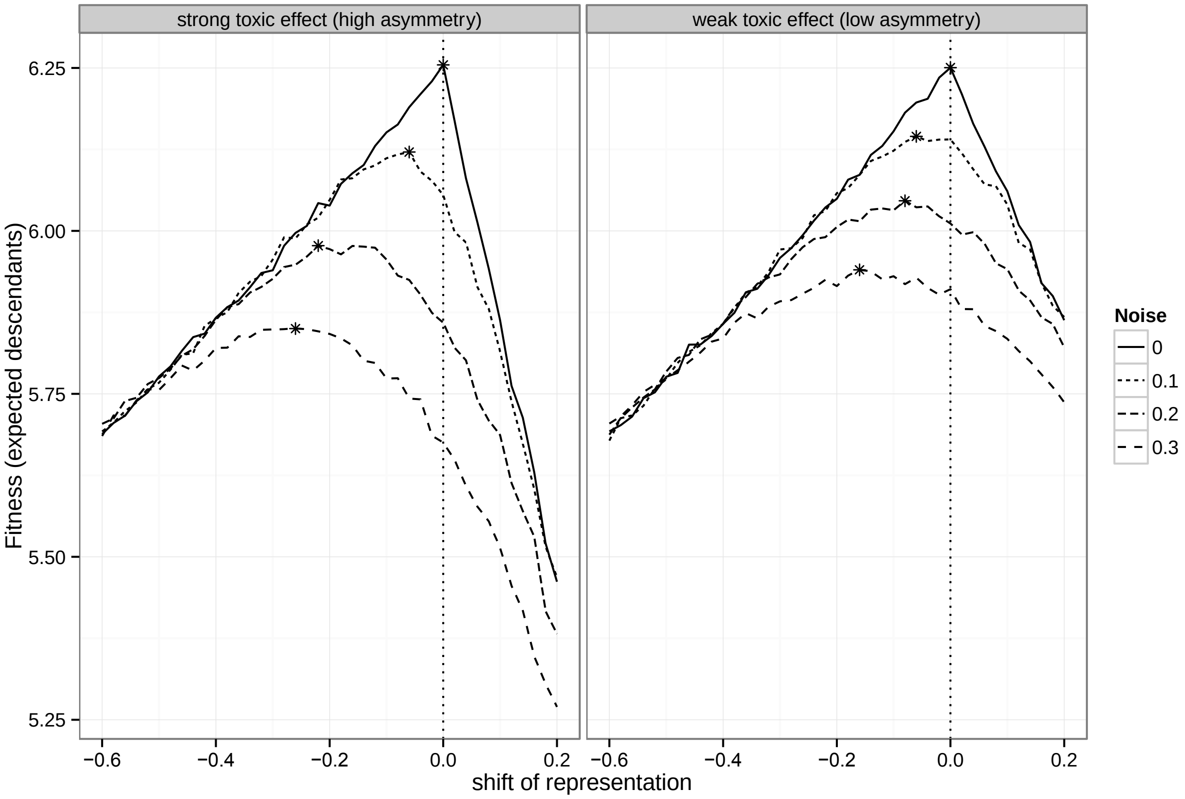


Figure 5. Average fitness of monkeys, depending on sensory accurateness (noise), shift of representation, and asymmetry of utility matrix (strong vs. weak toxic effect). The asterisk marks the optimal shift of representation for each noise level.

Given an asymmetric utility matrix and noise, fitness is optimized at some degree of misrepresentation < 0. Increasing noise levels are compensated by increasing shifts of representation.

Concerning the *truth* of a representation R_t, however, increasing levels of misrepresentation lead to a decreasing number of correct action, as can be seen in Figure 6. In contrast to the utility, the number of wrong decisions is independent of the utility matrix, and therefore symmetric around the zero point. With negative shift values, costly errors (missed toxic snakes) are reduced, but the less costly errors are disproportionately increased, leading to an increased overall level of false representations R_t.

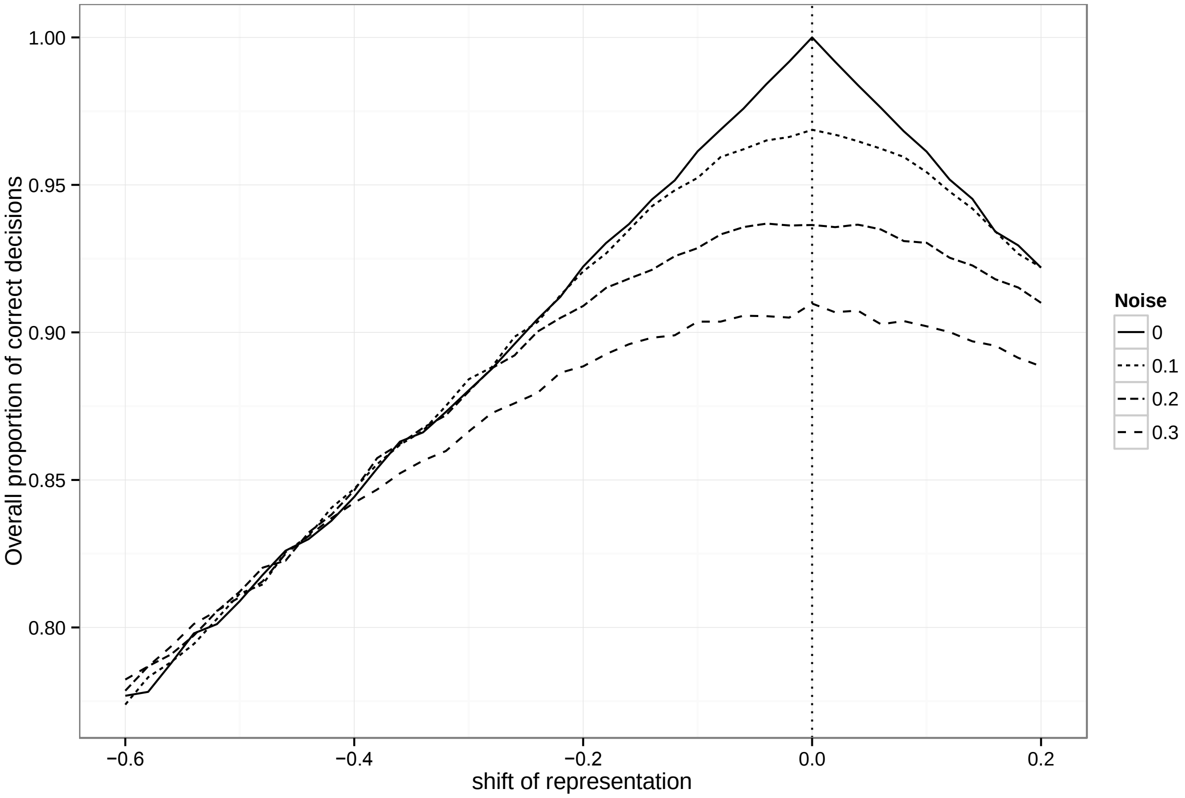


Figure 6. Overall number of correct decisions, depending on the amount of shift of representation.

In other words, if the representational shift is treated as the free parameter, one can predict that at least three factors lead to stronger misrepresentations: higher sensory noise, stronger asymmetry in the utility matrix, and a higher prior probability of encountering the more detrimental world state[12].

Discussion

First, we discuss the generality of our proposal, specifically pre-requisites and examples from illusions. Second, we discuss the relation to evolutionary epistemology, specifically as presented in Vollmer (1975) and Bischof (2009).

Generality

Omnipresence of indicator representations

[12] The last point is not shown in the simulation. Increasing the probability of encountering a toxic snake to values greater than 50% leads to lower overall levels of fitness and to a greater shift of representation. Probabilities smaller than 50% have the reverse effect.

The success of a false representation is not confined to the present example. Rather, it is a general phenomenon based on a generic principle. For the case without indicator variables (Figure 2), Godfrey-Smith (1991) and Bischof (1998) have demonstrated analytically based on signal detection theoretic (Green and Swets, 1966) and game theoretic (von Neumann & Morgenstern, 1947) considerations, respectively, that average utility can be maximised for systematically false representations. These false representations, however, in specific instances are in line with success semantics. However, when the success relevant feature has no direct causal impact on the organisms sensory system, indicator variables are used to form intermediate indicator representations, and false representations of the indicator variables can cause successful actions.

It seems to be rather easy to find examples, where indicator representations are used, i.e. $T_c \neq T_t$ (Figure 3) compared to finding cases, where the success relevant feature is directly sensed (Figure 2). Even in very simple organisms, for which the complete set of sensors, the set of behavioural repertoire, and most of the internal neuronal wiring is known, such as nematodes (C. elegans; Faumont, Lindsay, & Lockery, 2012), the sensory modalities such as touch to different parts of the organism's shape, chemotaxis (oxygen or carbon dioxide), or temperature, are not the success relevant features for different sets of behavioural actions, such as ingestion, defecation, feeding, or escape. The only example we were able to think of, where the success-relevant variable is identical with the indicator feature (i.e., T_t=T_c) is phototaxis in photosynthetic organisms (for a review see Jékely, 2009). Phototaxis is a "behavioral migration-response of an organism toward a change in illumination regime" (Hoff, van der Horst, Nudel, & Hellingwerf, 2009, p. 25). Positive phototaxis is a migration towards the light source, which is a successful action for photosynthetic organisms. It seems that apart from photosynthetic organisms, light sensors (such as eyes) rather generally produce indicator representations (similar to sound waves picked up by ears, or odours picked up by olfactory sensors).

The role of learning

Now, a large number of the indicator examples presented in the present chapter could be summarised as fixed action patterns (instincts) where a sign stimulus or releaser signal acts as an indicator to trigger a certain action which is adaptive for situations, which frequently correlate with the indicator in the environment of evolutionary adaptiveness (e.g., Tinbergen, 1951;

Lorenz, 1937; for a review see Schleidt, 1962). Beyond such fixed (innate) or acquired fixed action patterns, in associative learning in classical or operant conditioning, arbitrary coupling of any T_t with nearly any T_c (for limits see Breland & Breland, 1961) can be generated by an organism. For instance, drooling in dogs is an appropriate (successful) response to the presence of food, but ringing a bell (as an indicator feature; Pavlov, 1927) has no influence on drooling's success. In general, learning allows ontogenetic adaptation to the statistical properties of dynamically changing and unforeseen configurations of the environment. Thus, a flexible mechanism would possibly be implemented. In case misrepresentations in such learned contexts are optimal (i.e., in the presence of asymmetric priors or costs), misrepresentations should presumably happen between R_c to R_t. In the size version of our snakes example (Figure 4, right panel), a monkey could learn that small snakes are toxic and treat snakes as toxic up to a size of 25 cm, even though the true cut-off was 20 cm. These cases would be no challenge to success semantics.

However, the present chapter aims at demonstrating that there are cases, where there is little flexibility changes in R_c.

The case of illusions

The examples provided in this chapter mainly stem from the animal kingdom. Thus, two examples of misrepresentations are presented to demonstrate that systematic misrepresentations stably exist in humans for primary qualities (i.e., size and location), and can serve multiple action purposes: the Ebbinghaus size illusion and visual capture of sound.

In the Ebbinghaus (or Titchner) illusion, the relevant physical property is a disk's size *S* and a corresponding representation of size R_S in humans (see Figure 7). Specifically, the representation of size depends on the central disk's context: in case the central disk is surrounded by large circles, it is represented as being smaller and in case of small disks in the surround as being larger. The surround changes the size representation in the range of 5-10% (c.f. Franz, Gegenfurtner, Bülthoff, & Fahle, 2000).

The (mis-)representation of size, R_S, is a multipurpose representation that is causally involved in several different tasks/actions (Aglioti, DeSouza, & Goodale, 1995). Concerning perception, the context determines how big humans *see* the disk. When asked to *match* another disk to the size of the central disk ("perceptual matching"), humans over- or underestimate the disk size. The same happens when they are asked to *show* the disk size with their thumb and finger. And finally, when asking to *grasp* the central disk, the maximal grip aperture, that is the maximal opening of thumb and index finger when the hand is en route to the target, is also modulated by context (for a review see Franz, 2001). Physical size linearly affects maximal grip aperture, as well (Castiello, 2005). Finally, even though the wrong representation of size, R_S, has a causal influence on the grasping action, it is successful in picking up the central disk. Thus, it seems possible to successfully grasp a disk even though the representation of its size is false.

The second systematic misrepresentation in humans presented here concerns the primary quality location in space. Consider watching TV or a movie where a person speaks. Then the mouth has a certain location in space. Simultaneously with the lip movements, speech utterances are audible. What usually happens is, that the speech is perceived as originating from the speaker’s mouth, even though it actually originates from loudspeakers to the left and right of the screen. Here, there is a false representation of the sound’s origin in space, specifically, the sound’s origin is mislocalized to the location of the moving lips (which also allows ventriloquists to do their trick). In general terms, vision’s representation of spatial origin captures the auditory representation of spatial origin (e.g., Pick, Warren, & Hay, 1969; Warren, Welch, & McCarthy, 1981). Recently, it has been demonstrated, that sound can capture vision, if sensory noise in vision is increased (Alais & Burr, 2004) in such a way, that the integration of conflicting information about an event’s location of vision and space are combined optimally, depending on the levels of noise in both modalities (Ernst & Banks, 2002).

These example foremost serve the purpose to demonstrate that systematic misrepresentations are present in humans for primary qualities (such as size and location) even in situations where the respective representations are causally involved in a wider range of different types of actions. However, these examples serve also a second purpose: based on the conceptual framework of the present chapter, it is possible to ask, why such illusions (i.e., systematic misrepresentations) have stably evolved in human evolution. Remember, we could demonstrate that misrepresentations maximise fitness under certain conditions: (i) presence of sensory noise, (ii) asymmetry in priors or cost of errors, and possibly (iii) the presence of indicator representations. Consequently, the framework presented here provides a heuristic to ask, whether these prerequisites are met in the case of the presented illusions. Note, that it is possible that the Ebbinghaus or ventriloquist misrepresentations are not themselves maximising fitness but caused by a different organetic or other constraint.

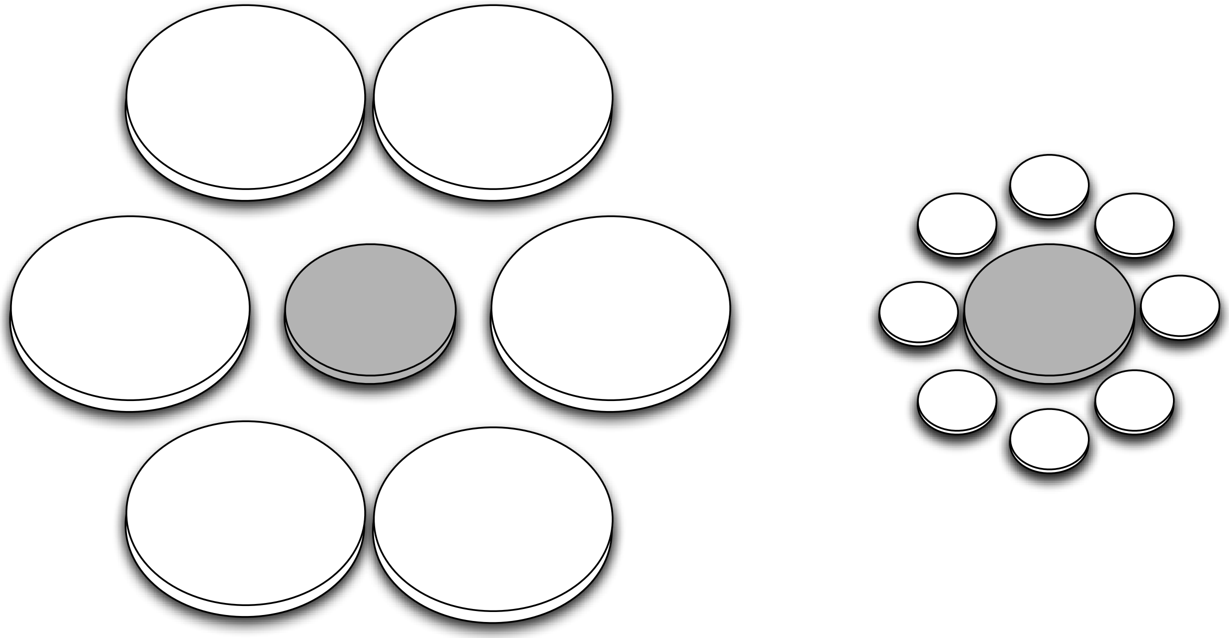


Figure 7. The central disk has (e.g., measured with a ruler) the same size when it is surrounded by large (left) or small (right) disks.

Evolutionary Epistemology

In the previous sections we argued that there can be situations where our representations systematically deviate from an objective reality. From a first-person point of view, however, it is not evident for a specific situation whether we misrepresent or not, as has been illustrated by several illusions. Given this situation, we can ask: What can be know about the world at all? Or, put in other words, What are the epistemological consequences from misrepresentations?

Evolutionary epistemology[13] (Lorenz, 1973; Vollmer, 1975) maintains the concept that our perception and representations have adapted to the (hypothesized) real world. From this general point of view several deductions can be made. On the one hand, our representations can be expected to be quite reliable and objective in domains which are highly fitness relevant. From the same theoretical framework one can also conclude that our representations will not be perfect: “In evolution, that is under competition, it pays to recognize outside objects more or less correctly. But it would not pay to aim at or to reach perfection.” (Vollmer, 2010, p. 1652).

[13] The term “evolutionary epistemology” has been used in at least two different notions (Bradie, 1986). Popper and others used the term to describe the growth of human knowledge by the (non-genetic) evolution of ideas and theories (Popper, 1972). In this chapter, we use it only in the sense of Lorenz (1973) and Vollmer (1975).

Finally, for domains which are not fitness relevant at all or no potentially successful actions can be performed by the organism, no selection pressure existed which would have shaped humans sensory abilities or representations towards objectivity.

Based on this general framework, several conclusions concerning misrepresentations have been drawn by Vollmer (1975) and Bischof (2009), as will be shown in the next paragraphs.

Micro-, meso-, and macrocosm

Domains of human knowledge which somehow relate to fitness-relevant domains of the external world can be shaped by evolution towards objectivity (Vollmer, 1975). There are, however, domains of knowledge which are completely unrelated to the external world (e.g., mathematical symbolic systems). There is no way to falsify or verify representations and beliefs within such symbolic systems by recurrence to the external world. Likewise, there are physical domains of the external world which have not been fitness-irrelevant in our history of evolution (e.g., strong atomic radiation). Without selection pressure, no detectors could have evolved for such physical domains, even if they have become fitness-relevant nowadays.

Fitness-relevant physical phenomena are predominantly located in a rather narrow range of physical scales: The retina is only sensitive to a small band of electromagnetic frequencies, size estimates below 0.5 mm and above some kilometres are nearly impossible, and time spans of nanoseconds or geological history are hard to imagine.

Based on this observation, Vollmer (1975) categorized the physical phenomena into three “cosms”, a) *microcosm*, which subsumes phenomena which are on a too small scale to be fitness relevant (e.g., sub-atomic structures), b) *mesocosm*, which describes phenomena on a medium scale, and c) *macrocosm*, which describes very large physical scales (e.g., cosmologic dimensions). Whenever knowledge domains exceed the mesocosmic scales to which our sensory apparatus and representational categories are adapted, these categories might be suboptimal or misleading. As a consequence, our intuitive sense of such phenomena can lead us astray. For example, it is hard to grasp the wave-particle dualism of electrons, even for experienced physicists.

To summarize, Vollmer’s framework of micro-, meso-, and macrocosm gives some guidelines where to expect objectivity in human perception, and in which domains our adapted senses and categories might be bad guides.

Meta-, para-, and orthocosm

Bischof shares the general framework of evolutionary epistemology with Vollmer, but he suggested an alternative classification of the “cosms”. According to Bischof (2009), it is not necessary to distinguish micro- from macrocosm. Both categories are completely irrelevant to fitness, and therefore can be combined into a single category which he calls *metacosm*. The metacosm describes all phenomena which are fitness-irrelevant, beyond their location on a physical scale. For example, the question of whether “Beauty and Truth are the same” is hardly fitness-relevant, and it can not be expected that encounters with the real world can give any evidence for or against that idea.

Furthermore, Bischof proposed to divide the mesocosm based on the symmetry of the utility surface (given that evolution does not favour “over-optimal” sensors, it is assumed that always some level of noise is present). The *orthocosm* describes all phenomena with a symmetric surface - representational errors on both sides of the ridge have more or less the same costs in terms of fitness. For these phenomena, representations can be assumed to be shaped towards truth, as objective representations are optimal. Although we never can be sure whether we have *reached* objectiveness (Vollmer, 2010), we can expect a convergence towards objectiveness. The *paracosm*, in contrast, denotes phenomena with asymmetric utility surfaces. Based on the arguments given above, in paracosm representations are expected to converge on a conservative level shifted away from objectivity, as this is optimal. Bischof locates most social categories in the realm of paracosm, whereas domains which require a physical interaction with the external world, like tool usage, predominantly are in orthocosm. For a graphical comparison of Vollmer’s (1975) and Bischof’s (2009) categories of evolutionary epistemology, see Figure 8.

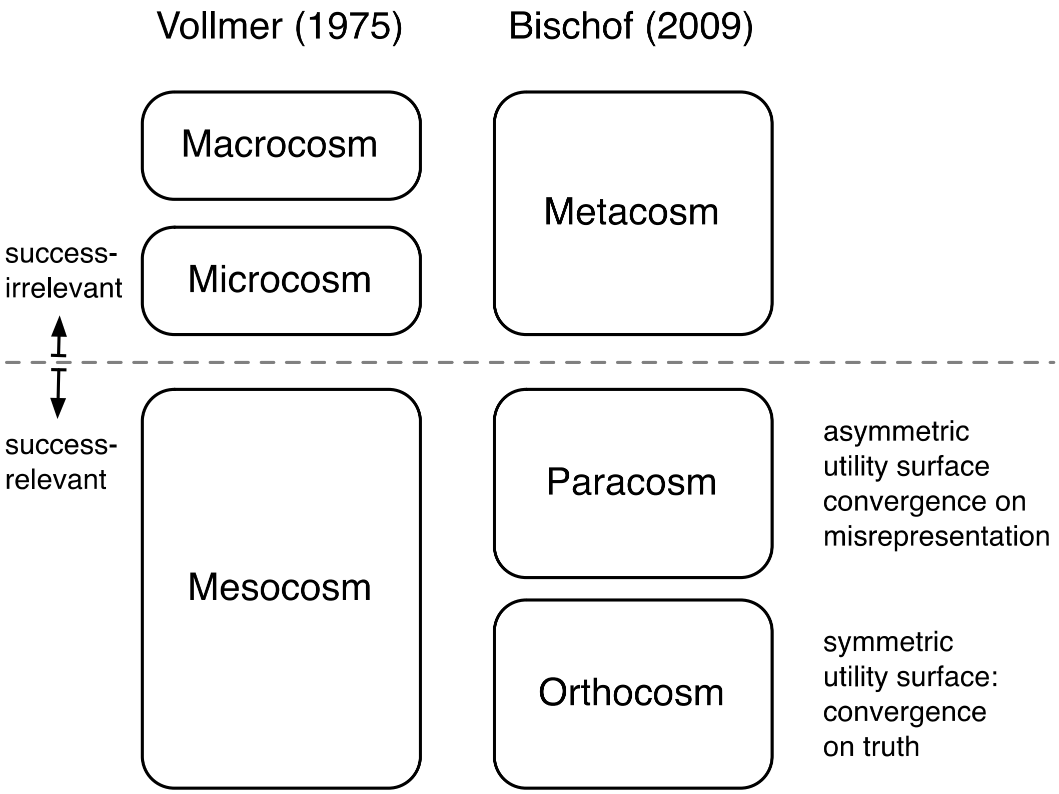


Figure 8. A graphical representation of Vollmer's (1975) and Bischof's (2009) epistemological categories.

Increasing objectivity: Multidimensional utility surfaces and measurement invariance

Both Vollmer and Bischof emphasize that multi-dimensional utility surfaces ("utility hyper surfaces") usually lead to more symmetry, and consequently to more objectivity. In our example of monkeys and snakes it is possible that a shifted colour perception (which is optimal for snake encounters) has adverse impacts in other fitness-relevant domains, for example "finding edible fruits". If this is the case, the overall utility hyper surface is a weighted average of all fitness-relevant tasks that make use of this particular sensor. This multi-dimensional utility hyper surface typically (but not necessarily) is more symmetric than the utility surface of a single dimension.

As evolution shaped our cognitive apparatus towards optimality and not towards objectivity, representations of paracosmic phenomena are not objective. But how can we, at least, approach objectivity for the paracosmic domain? Both Vollmer and Bischof agree that additional sensors can increase the objectivity of a representation. If multiple methods of measurement converge on the same result this would be a sign of *measurement invariance* (Bischof, 1966; Vollmer, 2010). Our monkeys have variance in colour perception in comparison to other measurement method:

Spectrometer measurements of the snake's colour, or the perception of other species that are immune to these snakes, would not converge with our monkey's perception. For Vollmer (2010), measurement invariance actually is the defining criterion for objectivity: "A proposition is objective if and only if its meaning and its truth is invariant against a change in the conditions under which it was formulated, that is, if it is independent of its author, observer, reference system, test method, and conventions" (p. 1658).

Hence, probing the colour perception of our monkeys with additional devices, independent of any toxicity associations, can be the tool to assess the objectivity of a representation (Vollmer, 2010). Although we never can be sure whether we have reached objectivity, invariance is the touchstone that tests for objectivity.

Conclusion

As has been argued before, under some conditions, namely information loss plus either asymmetry of the utility surface or prior probabilities, misrepresentations (or criterion shifts) lead to optimal actions. Success semantics state that a representation is true iff it causes a successful action. When misrepresentations are optimal, does that violate success semantics? Not necessarily. Although the average success is optimized for shifted representations, for each single action still holds that only true representations lead to successful actions.

In contrast to that situation, however, we argued that there are conditions in which false representations systematically lead to successful actions – not only in average, but also in single instances. Specifically, whenever (beyond information loss and asymmetry/ different priors) the organism employs indicator representations as proxies for the actual success relevant feature, systematically false representations in the indicator variable can lead to successful actions, and this situation indeed *is* a violation of success semantics. We provided examples hinting that the usage of indicator variables is probably more the rule than the exception in living organisms.

Embedding these ideas in the context of evolutionary epistemology, it can be assumed that humans have knowledge domains, which converge to the truth ("orthocosm"), because the overall utility hyper surfaces are symmetric. However, humans might also have knowledge domains where representations systematically deviate from the truth ("paracosm"), for asymmetric selection pressure.

References

Abrams, M. (2012). Measured, modeled, and causal conceptions of fitness. *Frontiers in Genetics, 3*, 1–12. doi:10.3389/fgene.2012.00196

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology: CB, 5*(6), 679–685.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology: CB, 14,* 257–262. doi:10.1016/j.cub.2004.01.029

Bazylinski, D. A., & Frankel, R. B. (2004). Magnetosome formation in prokaryotes. *Nature Reviews Microbiology, 2(3)*, 217–230. doi:10.1038/nrmicro842

Baube, C. L., Rowland, W. J., & Fowler, J. B. (1995). The mechanisms of colour-based mate choice in female threespine sticklebacks: hue, contrast and configurational cues. *Behaviour, 132(13-14)*, 13–14.

Bischof, N. (1966). Erkenntnistheoretische Grundlagenprobleme der Wahrnehmungspsychologie. In W. Metzger & H. Erke (Eds.), *Handbuch der Psychologie in 12 Bdn. Bd. 1/I: Wahrnehmung und Bewusstsein* (pp. 21–78). Göttingen: Verlag für Psychologie.

Bischof, N. (1998). *Struktur und Bedeutung*. Verlag Hans Huber.

Bischof, N. (2009). *Psychologie: Ein Grundkurs für Anspruchsvolle [Psychology: A basic course for the ambitious]* (2nd ed.). Kohlhammer.

Blackburn, S. (2005). Success semantics. In H. Lillehammer & D. H. Mellor (Eds.), *Ramsey's legacy* (pp. 22–36). Oxford: Oxford University Press.

Blakemore, R. P. (1975). Magnetotactic bacteria. *Science, 190(4212)*, 377–379.

Bradie, M. (1986). Assessing evolutionary epistemology. *Biology and Philosophy*, *1*, 401–459.

Breland K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist, 16,* 681-684.

Borchers, H.-W., Burghagen, H., & Ewert, J.-P. (1978). Key stimuli of prey for toads (Bufo bufo L.): Configuration and movement patterns. *Journal of Comparative Physiology, 128(3)*, 189–192.

Dretske, F. (1981), *Knowledge and the Flow of Information*. Cambridge, MA: MIT/Bradford Press.

Dretske, F. (1986), Misrepresentation. In R. Bogdan (ed.), *Belief* (pp. 17–36). Oxford: Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. doi:10.1038/415429a

Ewert, J.-P. (1974). The neural basis of visually guided behavior. *Scientific American*, *230(3)*, 34-42.

Faumont, S., Lindsay, T. H., & Lockery, S. R. (2012). Neuronal microcircuits for decision making in C. elegans. *Current Opinion in Neurobiology, 22(4)*, 580–591. doi:10.1016/j.conb.2012.05.005

Fodor, J. A. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fox, C. W., & Westneat, D. F. (2010). Adaptation. In D. F. Westneat & C. W. Fox (Eds.), *Evolutionary behavioral ecology* (pp. 16–32). New York: Oxford University Press.

Franz, V. H. (2001). Action does not resist visual illusions. *Trends in Cognitive Sciences, 5(11)*, 457–459.

Franz, V. H., Gegenfurtner, K. R., Bülthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: no evidence for a dissociation between perception and action. *Psychological Science, 11*, 20–25.

Green D.M., Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York Wiley.

Hoff, W. D., van der Horst, M. A., Nudel, C. B., & Hellingwerf, K. J. (2009). Prokaryotic phototaxis. *Methods in Molecular Biology, 571*, 25–49. doi:10.1007/978-1-60761-198-1_2

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in monkey striate cortex, *Journal of Comparative Neurology, 158(3),* 267-294.

Jékely, G. (2009). Evolution of phototaxis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1531)*, 2795–2808. doi:10.1098/rstb.2009.0072

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1968). What the frog's eye tells the frog's brain. In W. C. Cunning & M. Balaban (Eds.), *The mind: Biological approaches to its functions* (pp. 233–258). John Wiley & Sons.

Lorenz, K. (1937). Über die Bildung des Instinktbegriffes. *Die Naturwissenschaften, 25(19),* 289–300. doi:10.1007/BF01492648

Lorenz, K. (1973). *Die Rückseite des Spiegels: Versuch einer Naturgeschichte menschlichen Erkennens*. Munich, Germany: Piper.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy, 86(6),* 281–297.

Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton University.

Papineau, D. (1984). Representation and explanation. *Philosophy of Science*, *51*, 550–572.

Papineau, D. (2003). Is representation rife? *Ratio, 16(2),* 107–123.

Pavlov, I. P. (1927) in Pavlov, I. P. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences, 17(3),* 136–141. doi:10.5214/246

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*(4), 203–205. doi:10.3758/BF03207017

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach.* Oxford: Clarendon Press.

Ramsey, F. P., & Moore, G. E. (1927). Symposium: Facts and Propositions. *Proceedings of the Aristotelian Society, Supplementary Volumes, 7,* 153–206.

Reeve, H. K., & Sherman, P. W. (1993). Adaptation and the goals of evolutionary research. *Quarterly Review of Biology, 68*, 1–32.

Schleidt, W. M. 1962. Die historische Entwicklung der Begriffe "Angeborenes auslösendes Schema", und "Angeborener Auslösemechanismus" in der Ethologie. *Zeitschrift für Tierpsychologie, 19*, 697-722.

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science, 210(4471),* 801–803. doi:10.1126/science.7433999

Shea, N. (2007). Consumers need information: Supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research, 75(2),* 404–435.

Stephens, D. W., & Krebs, J. R. (1987). *Foraging Theory*. Princeton: Princeton University Press.

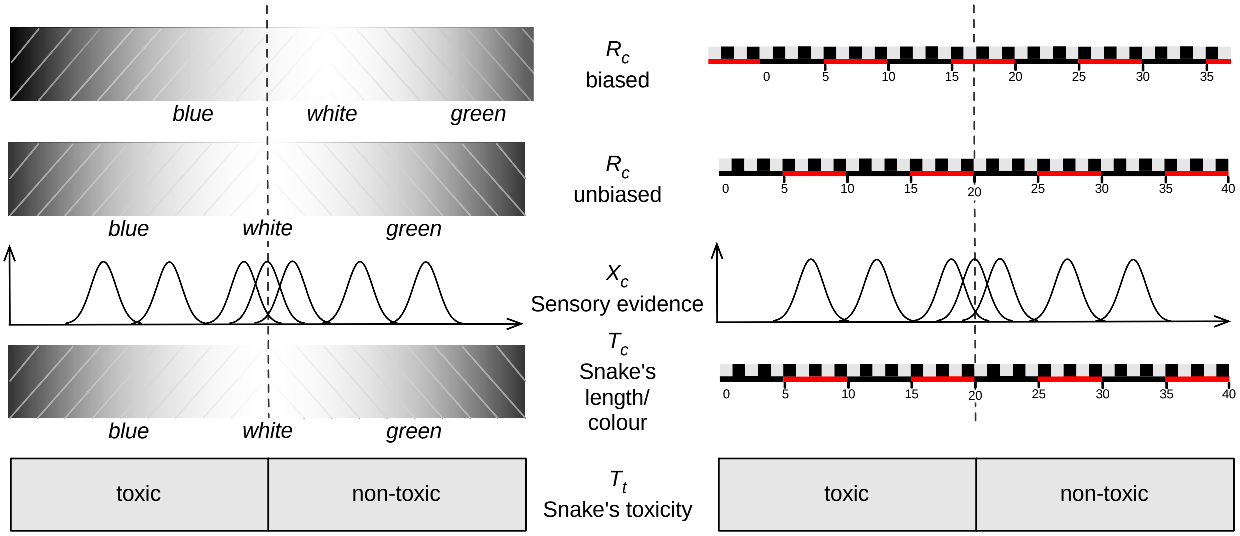
Tinbergen, N. (1951) *The study of instinct.* Clarendon Press, Oxford.

Vollmer, G. (1975). *Evolutionäre Erkenntnistheorie [Evolutionary epistemology]*. Stuttgart, Germany: Hirzel.

Vollmer, G. (2010). Invariance and objectivity. *Foundations of Physics, 40,* 1651–1667. doi:10.1007/s10701-010-9471-x

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564. doi:10.3758/BF03202010

Whyte, J. T. (1990). Success Semantics. *Analysis*, *50(3)*, 149–157.



Black-white version of Figure 4.