Science

AAAS

Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task

Petter Johansson *et al.*

Science **310**, 116 (2005);
DOI: 10.1126/science.1111709

This copy is for your personal, non-commercial use only.

to monitor baseline synaptic transmission (Fig. 4A). In wild-type slices, tetanic stimulation of pathway S1 (100 Hz, 1 s) evoked homosynaptic LTP together with a heterosynaptic depression of the neighboring S2 pathway. The addition of an A1 antagonist (DPCPX, 800 nM) prevented heterosynaptic depression (Fig. 4B). To control for effects of enhanced baseline transmission that result in the presence of DPCPX, we switched from normal artificial cerebrospinal fluid (ACSF) to one containing 2.4 mM Ca^{2+} and 0.6 mM Mg^{2+}, which enhanced synaptic transmission to $172 \pm 8.9\%$ ($n = 3$ slices) of that in control mice. We still found heterosynaptic depression to be intact ($64.9 \pm 8.2\%$, $n = 3$ slices) compared to ACSF controls ($72.0 \pm 8.6\%$, $n = 3$ slices). To determine whether astrocytes mediate adenosine-dependent depression, we repeated this study using dn-SNARE slices and found a virtual absence of heterosynaptic depression (Fig. 4C).

These studies place the astrocyte at center stage in the control of adenosine. Glial-released ATP, which is rapidly hydrolyzed to adenosine, leads to a persistent synaptic suppression mediated by A1 receptors. Because adenosine is implicated in the control of wake-to-sleep transitions (26, 27) as well as responses to hypoxia, the identification of the central role of the astrocyte in regulating this nucleoside offers mechanistic insights into these processes.

The kinetics of ATP hydrolysis and adenosine accumulation provide a synaptic network with unique spatiotemporal conditions to control synaptic transmission. Fast-acting synaptic transmitters such as γ-aminobutyric acid and glutamate have high-affinity uptake systems in the vicinity of the synapse that constrain the time and distance over which a transmitter acts. Synaptic activation of an astrocyte to release ATP removes these constraints, because it takes ~200 ms before adenosine begins to accumulate (28). This provides time for ATP diffusion to distant sites, where it depresses synaptic transmission through accumulated adenosine, thereby providing a mechanism for cross-talk to distant synapses. In addition to activity-dependent actions, astrocytes, by persistently suppressing excitatory synaptic transmission, enhance the capability of synapses to express synaptic plasticity. Thus, the integration of synaptic activity by the astrocyte leads to a widespread coordination of synaptic networks. By suppressing excitatory transmission, astrocytes regulate the degree to which a synapse may be plastic, and during the induction of LTP, astrocyte-derived adenosine depresses neighboring unstimulated pathways.

References and Notes

1. A. H. Cornell-Bell, S. M. Finkbeiner, M. S. Cooper, S. J. Smith, *Science* **247**, 470 (1990).
2. A. Verkhratsky, H. Kettenmann, *Trends Neurosci.* **19**, 346 (1996).
3. P. B. Guthrie et al., *J. Neurosci.* **19**, 520 (1999).
4. V. Parpura et al., *Nature* **369**, 744 (1994).
5. M. J. Schell, M. E. Molliver, S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3948 (1995).
6. T. Fellin et al., *Neuron* **43**, 729 (2004).
7. A. Araque, V. Parpura, R. P. Sanzgiri, P. G. Haydon, *Eur. J. Neurosci.* **10**, 2129 (1998).
8. J. Kang, L. Jiang, S. A. Goldman, M. Nedergaard, *Nat. Neurosci.* **1**, 683 (1998).
9. T. A. Fiacco, K. D. McCarthy, *J. Neurosci.* **24**, 722 (2004).
10. A. Araque, R. P. Sanzgiri, V. Parpura, P. G. Haydon, *J. Neurosci.* **18**, 6822 (1998).
11. A. Araque, N. Li, R. T. Doyle, P. G. Haydon, *J. Neurosci.* **20**, 666 (2000).
12. P. Bezzi et al., *Nat. Neurosci.* **7**, 613 (2004).
13. L. Pasti, M. Zonta, T. Pozzan, S. Vicini, G. Carmignoto, *J. Neurosci.* **21**, 477 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. Q. Zhang et al., *J. Biol. Chem.* **279**, 12724 (2004).
16. H. Zimmermann, N. Braun, *J. Auton. Pharmacol.* **16**, 397 (1996).
17. T. V. Dunwiddie, B. J. Hoffer, *Br. J. Pharmacol.* **69**, 59 (1980).
18. T. V. Dunwiddie, S. A. Masino, *Annu. Rev. Neurosci.* **24**, 31 (2001).
19. M. Kukley, M. Schwan, B. B. Fredholm, D. Dietrich, *J. Neurosci.* **25**, 2832 (2005).
20. K. A. Moore, R. A. Nicoll, D. Schmitz, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14397 (2003).
21. M. A. Ackley et al., *J. Physiol.* **548**, 507 (2003).
22. S. Coco et al., *J. Biol. Chem.* **278**, 1354 (2003).
23. O. J. Manzoni, T. Manabe, R. A. Nicoll, *Science* **265**, 2098 (1994).
24. J. T. Porter, K. D. McCarthy, *J. Neurosci.* **16**, 5073 (1996).
25. J. M. Zhang et al., *Neuron* **40**, 971 (2003).
26. T. Porkka-Heiskanen, L. Alanko, A. Kalinchuk, D. Stenberg, *Sleep Med. Rev.* **6**, 321 (2002).
27. R. Basheer, R. E. Strecker, M. M. Thakkar, R. W. McCarley, *Prog. Neurobiol.* **73**, 379 (2004).
28. T. V. Dunwiddie, L. Diao, W. R. Proctor, *J. Neurosci.* **17**, 7673 (1997).
29. We thank I. Levitan and T. Abel for constructive criticism on versions of this manuscript, T. Abel for his constant guidance, and R. Thresher, H. Bujard, and M. Brenner for constructs used to generate transgenic vectors. Supported by funds from the National Institute of Neurological Disorders and Stroke and the National Institute of Mental Health to K.M., S.J.M., and P.G.H.

Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task

Petter Johansson,[1]* Lars Hall,[1]*† Sverker Sikström,[1] Andreas Olsson[2]

A fundamental assumption of theories of decision-making is that we detect mismatches between intention and outcome, adjust our behavior in the face of error, and adapt to changing circumstances. Is this always the case? We investigated the relation between intention, choice, and introspection. Participants made choices between presented face pairs on the basis of attractiveness, while we covertly manipulated the relationship between choice and outcome that they experienced. Participants failed to notice conspicuous mismatches between their intended choice and the outcome they were presented with, while nevertheless offering introspectively derived reasons for why they chose the way they did. We call this effect choice blindness.

A fundamental assumption of theories of decision making is that intentions and outcomes form a tight loop (1). The ability to monitor and to compare the outcome of our choices with prior intentions and goals is seen to be critical for adaptive behavior (2–4). This type of cognitive control has been studied extensively, and it has been proposed that intentions work by way of forward models (5) that enable us to simulate the feedback from our choices and actions even before we execute them (6, 7).

However, in studies of cognitive control, the intentions are often tightly specified by the task at hand (8–10). Although important in itself, this type of research may not tell us much about natural environments where intentions are plentiful and obscure and where the actual need for monitoring is unknown. Despite all its shortcomings, the world is in many ways a forgiving place in which to implement our decisions. Mismatches between intention and outcome are surely possible, but when we reach for a bottle of beer, we very seldomly end up with a glass of milk in our hands. But what if the world were less forgiving? What if it instead conspired to create discrepancies between the choices we make and the feedback we get? Would we always

[1]Lund University Cognitive Science, Lund University, Kungshuset Lundagård, 222 22 Lund, Sweden. [2]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: lars.hall@lucs.lu.se

be able to tell if an error were made? And if not, what would we think, and what would we say?

To examine these questions, we created a choice experiment that permitted us to surreptitiously manipulate the relationship between choice and outcome that our participants experienced. We showed picture pairs of female faces to 120 participants (70 female) and asked them to choose which face in each pair they found most attractive. On some trials, immediately after their choice, they were asked to verbally describe the reasons for choosing the way they did. Unknown to the participants, on certain trials, a double-card ploy was used to covertly exchange one face for the other (Fig. 1). Thus, on these trials, the outcome of the choice became the opposite of what they intended. Each subject completed a sequence of 15 face pairs, three of which were manipulated (M). The M face pairs always appeared at the same position in the sequence, and for each of these pairs, participants were asked to state the reasons behind their choice. Verbal reports were also solicited for three trials of non-manipulated (NM) pairs (11).

The experiment employed a 3-by-2, between-group factorial design, with deliberation time and similarity of the face pairs as factors. For time, three choice conditions were included: one with 2 s of deliberation time, one with 5 s, and one where participants could take as much time as they liked. Participants generally feel that they are able to form an opinion given 2 s of deliberation time (supporting online text). Nevertheless, the opportunity for participants to enjoy free deliberation time was included to provide an individual criterion of choice. For similarity, we created two sets of target faces, a high-similarity (HS) and a low-similarity (LS) set (fig. S1). Using an interval scale from 1 to 10, where 1 represents "very dissimilar" and 10 "very similar," the HS set had a mean similarity of 5.7 (SD = 2.1) and the LS set a mean similarity of 3.4 (SD = 2.0).

Detection rates for the manipulated pictures were measured both concurrently, during the experimental task, and retrospectively, through a post-experimental interview (11) (supporting online text). There was a very low level of concurrent detection. With a total of 354 M trials performed, only 46 (13%) were detected concurrently. Not even when participants were given free deliberation time and a set of LS faces to judge were more than 27% of all trials detected this way. There were no significant differences in detection rate between the 2-s and 5-s viewing time conditions, but there was a higher detection rate in the free compared to the fixed viewing time conditions [$t(118)$ = 2.17, P = < 0.05]. Across all conditions, there were no differences in detection rate between the HS and the LS sets (Fig. 2A). In addition, there were no significant sex or age differences in detection rate. Tallying all forms of detec-

tion across all groups revealed that no more than 26% of all M trials were exposed.

However, these figures are inflated even so. The moment a detection is made, the outlook of the participants changes: They become suspicious, and more resources are diverted to monitoring and control. To avoid such cascading detection effects, it is necessary to discard all trials after the first detection is made. Figure 2B shows detection rates with this correction in place. The overall detection rate was significantly lower [$t(118)$ = 3.21, P < 0.005], but none of our prior conclusions are affected by the use of this data set (the percentage of participants that detected the manipulation is shown in fig. S2).

Our experiment indicates that the relationship between intentions and outcomes may sometimes be far looser than what current theorizing has suggested (6, 9). The detection rate was not influenced by the similarity of the face pairs, indicating the robustness of the finding. The face pairs of the LS set bore very little resemblance to each other, and it is hard to imagine how a choice between them could be confused (fig. S1 and supporting online text). The overall detection rate was higher when participants were given free deliberation time.

This shows the importance of allowing individual criteria to govern choice, but it is not likely to indicate a simple subjective threshold. The great majority of the participants in the 2-s groups believed themselves to have had enough time to make a choice (as determined by post-test interviews), and there was no difference in the actual distribution of choices among the pairs from fixed to free deliberation time.

Next, we examined the relationship between choice and introspective report. One might suspect that the reports given for NM and M trials would differ in many ways. After all, the former reports stem from a situation common to everyday life (revealing the reasons behind a choice), whereas the latter reports stem from a truly anomalous one (revealing the reasons behind a choice one manifestly did not make).

We classified the verbal reports into a number of different categories that potentially could differentiate between NM and M reports. For all classifications, we used three independent blind raters, and interrater reliability was consistently high (supporting online text and table S1). We found no differences in the number of empty reports (when participants were unable to present any reasons at all) or in the

Fig. 1. A snapshot sequence of the choice procedure during a manipulation trial. (**A**) Participants are shown two pictures of female faces and asked to choose which one they find most attractive. Unknown to the participants, a second card depicting the opposite face is concealed behind the visible alternatives. (**B**) Participants indicate their choice by pointing at the face they prefer the most. (**C**) The experimenter flips down the pictures and slides the hidden picture over to the participants, covering the previously shown picture with the sleeve of his moving arm. (**D**) Participants pick up the picture and are immediately asked to explain why they chose the way they did.
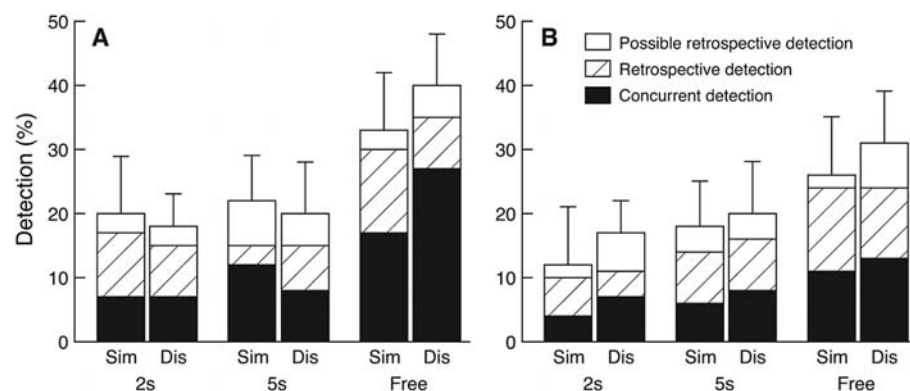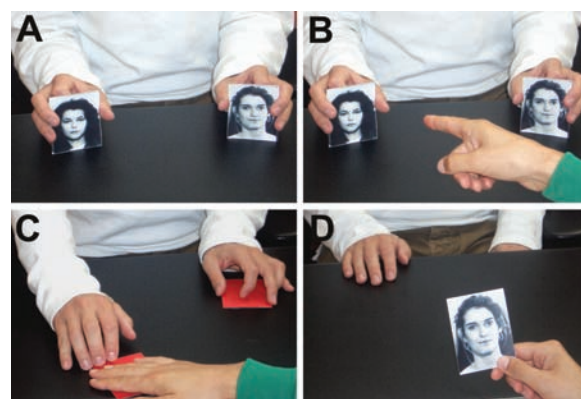
Fig. 2. Percent detection, divided into deliberation time and similarity, for (**A**) all trials and (**B**) trials corrected for prior detections. Sim, similar (HS); Dis, dissimilar (LS). Error bars indicate the standard deviation of the means.

degree to which reports were phrased in present or past tense (which might indicate whether the report is made in response to the present face or the prior context of choice). Neither did the length of the statements, as measured by number of characters, differ between the two sets (NM = 33, SD = 45.4; M = 38, SD = 44.4), nor the amount of laughter present in the reports (with laughter being a potential marker of nervousness or distress). We found significantly more dynamic self-commentary in the M reports [$t(118) = 3.31$, $P < 0.005$]. In this type of commentary, participants come to reflect upon their own choice (typically by questioning their own prior motives). However, even in the M trials, such reports occurred infrequently (5% of the M reports).

We rated the reports along three dimensions: emotionality, specificity, and certainty (using a numeric scale from 1 to 5). Emotionality was defined as the level of emotional engagement in the report, specificity as the level of detail in the description, and certainty as the level of confidence in their choice the participants expressed. There were no differences between the verbal reports elicited from NM and M trials with respect to these three categories (fig. S3). Seemingly, the M reports were delivered with the same confidence as the NM ones, and with the same level of detail and emotionality. One possible explanation is that overall engagement in the task was low, and this created a floor effect for both NM and M reports. However, this is unlikely to be the case. All three measures were rated around the midline on our scale (emotionality = 3.5, SD = 0.9; specificity = 3.1, SD = 1.2; certainty = 3.3, SD = 1.1). Another possibility is that the lack of differentiation between NM and M reports is an indication that delivering an M report came naturally to most of the participants in our task. On a radical reading of this view, a suspicion would be cast even on the NM reports. Confabulation could be seen to be the norm and truthful reporting something that needs to be argued for.

To scrutinize these possibilities more closely, we conducted a final analysis of the M reports, adding a contextual dimension to the classification previously used. Figure 3 shows the percentage of M reports falling into eight different categories. The "specific confabulation" category contains reports that refer to features unique to the face participants ended up with in a manipulated trial. As these reports cannot possibly be about the original choice (i.e., "I chose her [the blond woman] because she had dark hair"), this would indeed be an indisputable case of "telling more than we can know" (12). Equally interesting is the "original choice" category. These are reports that must be about the original choice, because they are inconsistent with the face participants ended up with (i.e., "I chose her because she smiled [said about the solemn one]". Here, despite the imposing context of the manipulated choice, vestiges of the original intention are revealed in the M reports. Analogous to the earlier example of confabulation, this would be an unquestionable case of truthful report.

In summary, when evaluating facial attractiveness, participants may fail to notice a radical change to the outcome of their choice. As an extension of the well-known phenomenon of change blindness (13), we call this effect choice blindness (supporting online text). This finding can be used as an instrument to estimate the representational detail of the decisions that humans make (14). We do not doubt that humans can form very specific and detailed prior intentions, but as the phenomenon of choice blindness demonstrates, this is not something that should be taken for granted in everyday decision tasks. Although the current experiment warrants no conclusions about the mechanisms behind this effect, we hope it will lead to an increased scrutiny of the concept of intention itself. As a strongly counterintuitive finding, choice blindness warns of the dangers of aligning the technical concept of intention too closely with common sense (15, 16).

In addition, we have presented a method for studying the relationship between choice and introspection. Classic studies of social psychology have shown that telling discrepancies between choice and introspection can sometimes be discerned in group-level response patterns (12) but not for each of the individuals at hand. In the current experiment, using choice blindness as a wedge, we were able to "get between" the decisions of the participants and the outcomes with which they were presented.

Type	%		
Specific Conf.	13.3		She's radiant. I would rather have approached her at a bar than the other one. I like earrings! [M]
Detailed Conf.	17.3	She looks like an aunt of mine I think, and she seems nicer than the other one. [F]	
Emotional Conf.	9.3		Yes, well, [laughter] she looks very hot in this picture. [M]
Simple Conf.	10.8		Just a nice shape of the face, and the chin. [M]
Relational Conf.	21.3		I thought she had more personality, in a way. She was the most appealing to me. [F]
Uncertainty	11.6	Eh... I don't know. [F]	
Dynamic report	5.2		Oh, [short laughter] Why did I choose her? She looks very masculine! [M]
Original choice	11.2	Because she was smiling. [F]	

Fig. 3. Frequency distribution of the contents of the M reports aligned along a rough continuum from confabulatory to truthful report. Sample sentences (translated from Swedish) are drawn from the set of reports for the displayed face pair. Letters in brackets indicate whether the report was given by a male (M) or a female (F) participant. The specific confabulation (Conf.) category contains reports that refer to features unique to the face participants ended up with in an M trial. The detailed and emotional confabulation categories contain reports that rank exceptionally high on detail and emotionality (>4.0 on a scale from 1 to 5). The simple and relational confabulation categories include reports where the generality of the face descriptions precluded us from conclusively associating them with either of the two faces (i.e., everybody has a nose, and a personality). The category of uncertainty contains reports dominated by uncertainty (<2 on a scale from 1 to 5). The dynamic reports are reports in which participants reflect upon their own choice, and the final category contains reports that refer to the original context of choice.

This allowed us to show, unequivocally, that normal participants may produce confabulatory reports when asked to describe the reasons behind their choices. More importantly, the current experiment contains a seed of systematicity for the study of choice and subjective report. The possibility of detailing the properties of confabulation that choice blindness affords could give researchers an increased foothold in the quest to understand the processes behind truthful report.

References and Notes
1. K. R. Ridderinkhof, W. P. M. van den Wildenberg, S. J. Segalowitz, C. S. Carter, *Brain Cogn.* **56**, 129 (2004).
2. K. R. Ridderinkhof, M. Ullsberger, E. A. Crone, S. Nieuwenhuis, *Science* **306**, 443 (2004).
3. M. Ullsperger, D. Y. Cramon, *Cortex* **40**, 593 (2004).
4. M. E. Walton, J. T. Devlin, M. F. S. Rushworth, *Nat. Neurosci.* **7**, 1259 (2004).
5. P. Haggard, S. Clark, *Conscious. Cogn.* **12**, 695 (2003).
6. R. Grush, *Behav. Brain Sci.* **27**, 377 (2004).
7. D. M. Wolpert, K. Doya, M. Kawato, *Philos. Trans. R. Soc. London Ser. B.* **358**, 593 (2003).
8. J. G. Kerns *et al.*, *Science* **303**, 1023 (2004).
9. R. Hester, C. Fassbender, H. Garavan, *Cereb. Cortex* **14**, 986 (2004).
10. H. C. Lau, R. D. Rogers, P. Haggard, R. E. Passingham, *Science* **303**, 1208 (2004).
11. Materials and methods are available as supporting material on *Science* Online.
12. R. E. Nisbett, T. D. Wilson, *Psychol. Rev.* **84**, 231 (1977).
13. R. A. Rensink, *Annu. Rev. Psychol.* **53**, 245 (2002).
14. D. J. Simons, R. A. Rensink, *Trends Cogn. Sci.* **9**, 16 (2005).
15. D. C. Dennett, *The Intentional Stance* (MIT Press, Cambridge, MA, 1987).
16. D. M. Wegner, *The Illusion of Conscious Will* (MIT Press, Cambridge, MA, 2003).
17. We thank D. de Léon, K. Holmqvist, P. Björne, P. Gärdenfors, T. Dickins, N. Humphrey, A. Marcel, Q. Rahman, E. Adams, N. Bolger, B. Cohen, and E. Phelps for useful comments and discussions. We also thank C. Balkenius for providing the illustrations for the article and P. Rosengren for invaluable advice concerning the use of card magic techniques. Supported by The New Society of Letters at Lund (P.J.), The Knowledge Foundation, The Erik Philip-Sörensen Foundation (L.H.), and the Swedish Research Council (S.S.).

Sexual Selection Can Resolve Sex-Linked Sexual Antagonism

Arianne Y. K. Albert* and Sarah P. Otto

Sexual selection is a potent evolutionary force. However, very few models have considered the evolution of female preferences for traits expressed in both sexes. Here we explore how female preferences coevolve with sexually antagonistic traits, which involve alleles that are beneficial to one sex but harmful to the other. We show that with a sexually antagonistic trait on the X chromosome (males XY, females XX), females evolve to prefer mates carrying alleles beneficial to daughters. In contrast, with a Z-linked trait (males ZZ, females ZW), females more often evolve mating preferences for mates carrying alleles beneficial to sons (that is, flashy displays).

Evolutionary biologists have long puzzled over how and why female preferences drive the evolution of exaggerated male traits. Generally, female preferences are thought to enhance a female's long-term fitness by increasing her offspring's fitness, either directly or through genetic associations between preference and trait loci (*1*, *2*). Nearly all models assume that females do not initially express the male display trait, or else they assume that the fitness effects are the same in males and females (*3*). However, traits subject to sexual selection will often be sexually antagonistic, for example, with "sexy" male traits benefiting males but reducing female fitness (*4*–*6*). If a trait increases male reproductive success at the cost of female viability, females must then choose between having attractive sons (and unfit daughters) and having ugly sons (but unencumbered daughters). As long as females can detect the genotypic differences among males at sexually antagonistic loci, we expect mating preferences to evolve as described by the models explored here.

Theory predicts that sexually antagonistic loci are more likely to remain polymorphic on the sex chromosomes (*4*, *5*). Furthermore, recent empirical work suggests that many sexually selected traits in animals are located on the X chromosome (*7*, *8*) and that most polymorphic sexually antagonistic traits are located on the X chromosome in *Drosophila* (*5*, *9*). There is also evidence to suggest that a female's mate choice may result in a tradeoff in the fitness between her daughters and sons (*10*). Several recent theoretical examinations of the evolution of female preferences have explored sex linkage of the trait and/or preference (*11*–*14*). However, these models, with the exception of Reeve and Pfennig's (*12*),

assume that sexually selected traits have male-limited expression and therefore no fitness consequences when in females, and none has addressed sexual antagonism. Here we address the question of how female preferences evolve for traits that have contrasting fitness effects in each sex.

With sexual antagonism, chromosomal location should strongly affect the evolution of female preferences. Simply put, an X-linked male trait is never passed on from an attractive father to his sons, whereas his daughters suffer the cost of carrying the display trait (*5*, *9*). Offspring in XY species therefore do not gain a fitness benefit from females preferring males with a more extreme X-linked display trait. In contrast, both males and females contribute a Z chromosome to sons in ZW species. Thus, females preferring a Z-linked display trait receive the fitness benefit of sexy sons, even though their daughters suffer a fitness cost (*5*, *9*). This cost is lessened by the fact that daughters inherit only one of their father's Z chromosomes. With autosomal inheritance, these asymmetries in inheritance are absent.

To verify the verbal argument laid out above, we present the results of two-locus models that follow the fate of a newly arisen preference allele p in a population that is at a polymorphic equilibrium at a trait locus. We assume that

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: albert@zoology.ubc.ca

Table 1. Male and female fitness components in male heterogametic (XY) and female heterogametic (ZW) species.

		X-linked trait			Z-linked trait		
		Male trait					
		T	t		TT	Tt	tt
Female preference	PP	1	$1 + a_{PP}$	P	1	$1 + da_P$	$1 + a_P$
	Pp	1	$1 + a_{Pp}$	p	1	$1 + da_P$	$1 + a_P$
	pp	1	$1 + a_{pp}$				
Male viability		$1 - s_y$	1		$1 - s_z$	$1 - hs_z$	1
		Female trait					
		TT	Tt	tt		T	t
Female viability		1	$1 - hs_x$	$1 - s_x$		1	$1 - s_w$
		Female preference					
		PP	Pp	pp		P	p
Female cost		$1 - |a_{PP}|k$	$1 - |a_{Pp}|k$	$1 - |a_{pp}|k$		$1 - |a_P|k$	$1 - |a_P|k$