Unraveling the true complexity of costly color signaling

Darrell J. Kemp, Marie E. Herberstein, and Gregory F. Grether

Conspicuous displays of color comprise an enormously diverse and functionally complex class of biological signals. Many of these displays are widely publicized as resulting from chemical colorants known as pigments, which act by selectively “absorbing” part of the light spectrum (Appendix 1). However, the full diversity of animal coloration is just as strongly influenced by optically active surface structures, which act by selectively “reflecting” light (i.e., scattering). In cases where light is scattered coherently, these structural colors produce displays that are often metallic, iridescent, or kaleidoscopic in appearance. In addition to structural coloration, mechanisms such as bioluminescence and fluorescence allow the actual creation of colored light, thereby transcending the simple reflectance of whatever ambient wavelengths are available. There is enormous variation among all of these classes of colorants, and most signals arise through the interaction of diverse suites of pigments and reflecting (structural) mechanisms. Despite this complexity, however, the field of behavioral ecology has been largely preoccupied with understanding how carotenoids—a single group of pigments—may broker the honest signaling of individual quality. This focus has proven highly productive in many respects, but it has also fostered a limited and overly simplistic view of how color signals really work. As we outline here, the complete emerging story of animal coloration is far more complex and interesting. [Behav Ecol 23:233–236 (2012)]

“CAROTENO-CENTRICITY” CHALLENGED

Grounded in the theory of honest signaling (Zahavi 1975), the “caroteno-centric” focus on color signals began in the 1980s and rose to fruition over the next 2 decades. This paradigm had its inception in a simple, clearly articulated proximate connection between color expression and resource acquisition (Endler 1980). Carotenoids are somewhat unique pigments in that they cannot be synthesized de novo by most animals (although see Moran and Jarvik 2010) and must instead be acquired directly from the environment. The costs or difficulties of achieving this are thought, in turn, to predispose the resultant visual signals as revealing indicators of individual quality (Olson and Owens 1998). In this way, individuals looking to choose among potential mates could use carotenoid-based coloration as a proxy for mate quality. The honesty of carotenoid signals is also thought to be mediated through tradeoffs with other metabolically important demands, such as antioxidant and/or immune function (for an alternative viewpoint, see Hartley and Kennedy 2004).

Decades of research across birds, insects, fishes, and reptiles have delivered voluminous support for the role of carotenoid signals as honest condition-dependent mating signals (Olson and Owens 1998). Not all data are supportive, however (Griffith et al. 2006; see below), and the proximate pathways of signal expression have turned out to be considerably more complex than originally imagined (e.g., Grether et al. 2001; Hartley and Kennedy 2004). In guppies, for example, we now know that their orange dermal coloration arises not only from carotenoid pigments (as classically noted by Endler 1980) but also from the presence of pteridines—a completely different class of pigmentation (Grether et al. 2001). Pteridines are relatively nitrogen rich, but they are readily synthesized de novo and are subsequently thought to be relatively cheap to express. Indeed, because they can arise as by-products of other metabolic processes, an early view was that pteridines were essentially “excretory substances” that were subsequently co-opted for use in color signaling (Hopkins 1895). Interestingly, Grether et al. (2005) found that guppy populations subject to different environmental carotenoid availabilities differ genetically in their use of pteridines to color their orange spots, which is an important sexual trait (Endler and Houde 1995). The interaction between these 2 pigment suites is such that orange hue (ca. color) is held relatively constant along a carotenoid availability gradient (Grether et al. 2005). Such mechanistic complexity was unheralded several decades ago and at least calls for a reappraisal of the relative contribution of carotenoids to signal honesty in this system.

Although the costly carotenoid paradigm has fueled productive empirical research programs, particularly in showcase avian taxa (e.g., house finches and great tits; Hill and Montgomerie 1994; Horak et al. 2001), recent years have heralded a conceptually broader approach to interpreting color-based ornaments. This has come about primarily due to a more thoughtful appraisal of the many and diverse ways in which these traits may be costly or difficult to express (Griffith et al. 2006). Whereas one line of argumentation emphasized
the challenges of obtaining and appropriating carotenoids (e.g., McGraw et al. 2002), researchers have realized that the very same acquisition-based arguments could apply to the materials needed to assemble synthetic pigments, such as melamins (Griffith et al. 2006; Stoehr 2006). These arguments likewise apply to the cuticular or dermal materials needed to construct the light-scattering surface arrays responsible for structural coloration. Furthermore, given a broader, more taxonomically balanced view of viability, it is clear that acquisitional limitations offer just one of many potential proximate routes to signal honesty. Any reallocation of nutrients, energy, time, or any other important somatic resource to sexual advertisement could be treated as a viability handicap. Indeed, the aggregate of all such investments has been formalized as the true cost of sexual trait expression (Rowe and Houle 1996; Bonduriansky 2007a, 2007b). Even broader lines of argument propose that honesty in color signals can be forced by the sheer difficulty of producing particular colorants, such as the developmental challenges associated with assembling (or guiding the self-assembly of) the nanoscale elements of structural coloration (Fitzpatrick 1998; Kemp and Rutowski 2007). These more inclusive views of signal honesty are being increasingly explored, but we still know relatively little of their role in the evolution and diversity of animal coloration.

BROADENING THE SPECTRUM

Apart from allowing greater insights into particular systems (e.g., guppies), this broadened view of mechanisms, costs, and constraints is driving a paradigm shift in the biological appraisal of coloration. Not only has this viewpoint allowed researchers to generate and test new hypotheses regarding the proximate pathways of signal reliability (e.g., Fitzpatrick 1998; Kemp and Rutowski 2007), it has forced a greater understanding of the optics of color itself (Appendix 1) and of the diversity of ways in which it may be produced (Prum et al. 1998; Grether et al. 2004; Ghiradella 2010; Thery and Gomez 2010). Whereas entire classes of nonpigment colorants—such as the enormously varied phenomenon of structural coloration (Fitzpatrick 1998; Kemp and Rutowski 2007)—were previously considered as relatively invariant, "cheap" to produce—and hence—uninformative (although see McGraw et al. 2002), an increasing body of work is demonstrating precisely the opposite (Doucet and Meadows 2009). It is interesting to note that the very same paper that sparked initial interest in carotenoids (Endler 1980) also discussed the potential costliness of expressing extremely bright "structural," particular color signals are being increasingly understood to result from the composite effects of both absorbing pigments and reflective structures (Grether et al. 2004; Rutowski et al. 2005; Shawkey and Hill 2005). Studies targeting the roles of individual signal mechanisms have revealed interesting interactions in the way they contribute to overall signal design and appearance. In male Colias eurytheme butterflies, for example, the presence of underlying pteridine pigments has been found to amplify the iridescence of bright ultraviolet (UV), a structural color (Rutowski et al. 2005). Due to several architectural features of the nanostructures located on the wing scales which cause the coherent scattering of UV (for optical categorization of these structures, see Vukusic et al. 2001), this component of wing coloration is extremely limited, being visible for only a ~15° range of above-wing viewing angles. However, because the full wing sweep of a flying male C. eurytheme spans >100° (Rutowski et al. 2007), a stationary UV-sensitive observer would see a brief periodic “flash” of bright UV from the beating wing, contributing to a significant hue and brightness shift from its otherwise yellow/orange appearance. As Rutowski et al. (2005) discovered, the pteridine pigments accentuate the magnitude of this iridescent flash (i.e., the difference between the “on” and “off” positions for seeing UV) by extinguishing all the short-wave light that would otherwise scatter from the wing ultrastructure when in the off position for seeing the coherently scattered UV. This ensures that the wing is maximally UV-dark when in the off position, thereby maximizing contrast with the extreme UV-brightness of the wing when in the on position. Clearly, in these situations, focusing on just one aspect of the color signal is likely to underestimate (sometimes grossly) the true dynamic nature of how that signal appears, and influences fitness, in realistic biological contexts.

A BRIGHT FUTURE

The ultimate challenge for research on color signaling is to understand how the diverse color-producing mechanisms interact over evolutionary time. This will require an understanding of how the different components of color signals work together—not only in terms of their combined effects on light absorption and transmission but also in terms of their combined effects on signal efficacy and honesty. For example, in the case of color patches that result from different suites of pigments (such as the carotenoids and pteridines that underpin guppy oranges, Grether et al. 2005), what is the functional relationship between these light-absorbing elements? Is one merely a cheaper substitute for the other or are they both required for maximal signal efficacy? We could also ask whether both types of pigments function as indicators and, if so, are they redundant signals or do they correlate with different aspects of signaler quality? Similar questions should be posed regarding the functional relationship between light-absorbing and light-reflecting elements of individual color patches. For example, do reflecting elements function to amplify differences in pigment content (Grether et al. 2004) or do the pigments function to amplify differences in reflecting layer ultrastructures (e.g., Rutowski et al. 2005, as above)?

Empirically, what is required is a far more comprehensive and rigorous appraisal of the sources of variation in color—both at the phenotypic and genetic level—and at a resolution appropriate to individual mechanistic components (i.e., the nature and visual consequences of variation in a given pigment or a given surface architecture). The tools of quantitative genetics will be required to examine genetic variances, adaptive potential, and genetic covariances (e.g., with organismal viability traits, etc.) and to ultimately test hypotheses relating to signal evolution. Artificial selection experiments pose a powerful and direct approach to testing many of these questions. Such experiments could be used to explore how the different components of color signals respond to selection on specific spectral properties, thereby allowing the evaluation of mechanistic models for how structures, pigments, and cell layers interact to produce different colors. This approach could also be used to rigorously test explicit evolutionary scenarios, for example, the hypothesis that selection on orange spot hue is sufficient to explain why guppies in high-carotenoid availability environments deposit more pteridine pigments in their orange spots than do guppies in low-carotenoid availability environments (Grether et al. 2001, 2005). Finally, selection experiments allow opportunities for assessing the correlated phenotypic responses to selection and could thereby be used to yield insights into the quantitative genetic architecture of color traits and of their individual mechanistic and visual elements.
Conceptually, the idea that the production costs of carotenoids (or any other single component) are sufficient to keep color signals honest ought to be viewed with skepticism. We need to look beneath the surface, literally, and ask: how could the honesty of a color signal be maintained over evolutionary time if there are cheaper ways of producing it? If carotenoids are particularly costly, why have they not been replaced with other types of pigments or structural components that produce similar colors? One possible answer is that signaler–receiver evolution has resulted in highly discriminating receivers that are not fooled by cheaper substitutes. Another possibility is that color signals are already as cheap to produce as they could be and that alternative mechanisms of producing the same colors are even costlier. Testing either of these hypotheses would require a detailed understanding of the proximate mechanisms of color production and the available alternatives. Production costs are notoriously difficult to measure and compare directly, but the hypothesis about receiver discrimination should be testable. Mathematical color patch models, parameterized for the study system (e.g., Grether et al. 2004), could be used to examine the effects on the color produced of varying each color patch component. The hypothesis that receivers are so discriminating that cheating is no longer possible could be tested with behavioral experiments (e.g., mate choice tests) and phenotype manipulations that mimic or replicate, in vivo, the possible color variations. If receivers were indeed fooled by, say, a change in a structural component, this would put the research focus on understanding why this change has not occurred in natural populations (the answer might involve production costs, lack of genetic variation, etc.). We do not mean to downplay the difficulty of testing these hypotheses; our main point is simply that it is not logical for research on color signal honesty to focus exclusively on carotenoids or on any other single color patch component.

FUNDING
D.J.K. was supported by Macquarie University internal (startup) funds (9200900077) and Safety Net grant (9200901692).

We thank Robert Brooks and 1 anonymous reviewer for helpful comments on this manuscript. We also thank Kevin Clarke (of http://bugunderglass.com) for kindly providing the butterfly wing images.

APPENDIX 1: COLOR PRODUCTION IN NATURE

Color in nature can be produced by 4 principle but "non-mutually exclusive" mechanisms:

1. Pigmentary absorption (Panel A): The presence of biochemical compounds (i.e., pigments) in surface layers that selectively absorb part of the light spectrum. Ommochrome pigments, for example, which strongly absorb short wavelengths (i.e., ca. <500 nm), produce the "yellowish" dorsal band markings of the spider Gas- teracantha fornicata (Araeaceae).

2. Structural reflectance (Panel B): The presence of optically functional surface or surface layer architectures that selectively reflect part of the light spectrum. Multi-layer surface nanoarchitectures, for example, produce the green wing markings of the butterfly Papilio palinurus (Papilionidae).

3. Fluorescence (Panel C): The presence of biochemical compounds that absorb (usually selectively) part of the light wavelength spectrum and reemit the energy in a different region of the spectrum. Fluorescent markings in the budgerigar Melopsittacus undulatus (Psittacidae), for example, arise from underlying pigments which absorb in the UV region and which reemit light in long wave-lengths. This produces the human "yellow" markings seen in the right-most image, which is illuminated by UV light only (Arnold et al. 2002).


REFERENCES


